



Université de Lille, Faculté des sciences et technologies

École doctorale de Biologie Santé de Lille

THESE DE DOCTORAT

Caractérisation structurale et fonctionnelle de PII1, une protéine impliquée dans la régulation de l'initiation de la biosynthèse des grains d'amidon chez la plante

En vue de l'obtention du titre de Docteur mention aspects moléculaires et cellulaires de la biologie

Présentée par

Mélanie Bossu

Le 2 décembre 2024 devant la commission d'examen composée de :

Dr Tony Lefebvre	Pr, Université de Lille	Président
Dr Véronique Brechot	DR, CNRS, Université de Marseille	Rapportrice
Dr Mirjam Czjzek	DR, CNRS, Université Sorbonne, Roscoff	Rapportrice
Dr Arnaud Gruez	MCf, Université de Lorraine	Examinateur
Dr Julie Bouckaert	DR, CNRS, Université de Lille	Examinatrice
Dr Coralie Bompard	CR, CNRS, Université de Lille	Directrice de thèse

Unité de Glycobiologie Structurale et Fonctionnelle - UGSF

Villeneuve d'Ascq

Résumé

L'amidon est un polymère de glucose synthétisé par les plantes chez lesquelles il représente la principale forme de stockage d'énergie. Il est localisé dans les plastes sous forme de grains insolubles. L'amidon est composé principalement de deux polyglucanes, l'amylose, formé de chaînes linéaires et l'amylopectine, plus ramifiée, s'assemblent pour former le grain semi-cristallin. La synthèse de l'amidon implique l'action coordonnée de plusieurs enzymes, notamment les amidon synthases (SS), les enzymes de branchement (BE) et de débranchement (DBE). Parmi ces enzymes, la SS4 joue un rôle primordial dans l'initiation de la synthèse de l'amidon. SS4 est responsable de la régulation du nombre de grains d'amidon et de leur forme dans les chloroplastes. Chez *Arabidopsis thaliana*, les plantes mutantes dépourvues de SS4 produisent moins de grains par chloroplaste, et ces derniers sont plus gros et ronds par rapport aux grains lenticulaires observés chez les plantes sauvages. SS4 a été décrite comme étant l'enzyme principale impliquée dans l'initiation de la biosynthèse des grains d'amidon.

Dans la plante, SS4 interagit avec d'autres protéines non enzymatiques pour former un complexe d'initiation. Parmi celles-ci, PII1 a été décrite comme impliquée également dans l'initiation. Les mutants dépourvus de *pii1* présentent un phénotype similaire à celui des mutants SS4, avec un seul gros granule lenticulaire par chloroplaste, bien que ce phénotype soit moins sévère que celui du mutant ss4, qui présente des chloroplastes vides. Ces résultats suggèrent que PII1 est importante pour une initiation normale de la synthèse des grains d'amidon. PII1 fait partie du complexe d'initiation comprenant SS4, PTST2, PTST3, MFP1 et SS5. Toutes ces protéines contiennent des régions coiled-coils prédites suggérant des interactions protéine/protéine. PII1 ne possède ni activité catalytique ni domaine de fixation aux glucanes, ce qui suggère un rôle plutôt mécanique ou de régulation et probablement une implication dans l'initiation à travers l'interaction qu'elle forme avec SS4, PTST2 et SS5. Il a été proposé que PII1 agisse spécifiquement sur la capacité de SS4 à contrôler le nombre d'évènements d'initiation de l'amidon et pourrait modifier sa conformation ou plus largement modifier la stabilité d'un complexe protéique incluant SS4 et d'autres facteurs du complexe d'initiation.

Ce projet de recherche vise à caractériser la fonction moléculaire et fonctionnelle de PII1 et de SS4 en utilisant des approches structurales biophysiques et biochimiques. L'objectif plus global est d'avancer dans la compréhension du mécanisme d'initiation de la biosynthèse de l'amidon. Les protéines PII1 et SS4 ont été exprimées chez *E. coli*. Des constructions tronquées ont été synthétisées afin de faciliter la purification et utilisées pour les études d'activités enzymatiques et analyses biophysiques, dont le dichroïsme circulaire. En parallèle, des analyses bio-informatiques ont été menées pour modéliser la structure des protéines seules et en complexe. Les résultats obtenus ont permis de décrire la structure de PII1 et de montrer qu'elle est capable d'inhiber l'activité synthase de SS4 et d'avancer dans la compréhension de son rôle dans les mécanismes moléculaires de l'initiation de la biosynthèse de l'amidon chez la plante.

Abstract

Starch is a glucose polymer synthesised by plants, in which it is the main form of energy storage. It is found in the plastids in the form of insoluble granules. Starch is mainly composed of two polyglucans, amylose, which is made up of linear chains, and amylopectin, which is more branched, which combine to form the semicrystalline grain. Starch synthesis involves the coordinated action of several enzymes, in particular starch synthases (SS), branching enzymes (BE) and debranching enzymes (DBE). Among these enzymes, SS4 plays a key role in initiating starch synthesis. SS4 is responsible for regulating the number and shape of starch granules in chloroplasts. In Arabidopsis thaliana, mutant plants lacking SS4 produce fewer grains per chloroplast, which are larger and rounder than the lenticular grains observed in wild-type plants.

In plants, SS4 interacts with other non-enzymatic proteins to form an initiation complex. Among these, PII1 has been described to be involved in initiation. Mutants lacking PII1 show a phenotype similar to that of SS4 mutants, with a single large lenticular granule per chloroplast, although this phenotype is less severe than that of SS4 mutants, which show empty chloroplasts. These results suggest that PII1 is important for the normal initiation of starch granule synthesis. PII1 is part of the initiation complex that includes SS4, PTST2, PTST3, MFP1 and SS5. All these proteins contain predicted coiled-coil regions suggesting protein-protein interactions. PII1 has neither catalytic activity nor a glucan-binding domain, suggesting a mechanical or regulatory role and probably an involvement in initiation through the interaction it forms with SS4, PTST2 and SS5. It has been proposed that PII1 acts specifically on the ability of SS4 to control the number of starch initiation events and could alter its conformation or, more broadly, the stability of a protein complex including SS4 and other factors in the initiation complex.

This research project aims to characterise the molecular and functional function of PII1 and SS4 using biophysical and biochemical structural approaches. The overall aim is to improve our understanding of the initiation mechanism of starch biosynthesis. PII1 and SS4 proteins were expressed in E. coli. Truncated constructs were synthesised to facilitate purification and used for enzymatic activity studies and biophysical analyses, including circular dichroism. In parallel, bioinformatics analyses were carried out to model the structure of the proteins alone and in complex. The results obtained have enabled us to describe the structure of PII1 and to show that it is capable of inhibiting the synthase activity of SS4, thereby advancing our understanding of its role in the molecular mechanisms involved in the initiation of starch biosynthesis in plants.

Remerciements

Je tiens à adresser mes premiers remerciements à ma directrice de thèse, Dr Coralie Bompard. Ce fut un véritable plaisir de travailler sur ce projet de thèse passionnant. Je te remercie sincèrement pour ton expertise dans des domaines de pointe et pour ta rigueur scientifique, qui ont grandement contribué à l'amélioration de mon travail. Ton soutien sans faille à Rayan et moi, ainsi que le temps que tu nous as consacré, ont toujours été pour moi une source précieuse de motivation. Je te suis également reconnaissante pour les nombreuses opportunités, tant scientifiques que personnelles, que tu m'as offerte : congrès de Biophysique à Marseille, sessions au Synchrotron à Paris, workshop de biologie structurale à Oléron, médiation scientifique, Pauses thé etc. Grâce à toi, j'ai pleinement profité de ces années de doctorat et j'ai eu la chance de rencontrer des chercheurs passionnés et inspirants. Mille mercis pour ces années exceptionnelles qui resteront gravées dans ma mémoire.

Je souhaite également remercier chaleureusement Dr David Dauvillée, mon mentor en biologie moléculaire. Je ne te remercierai jamais assez pour tes innombrables conseils pertinents et avisés, qui ont largement contribué à l'avancement de mon travail de thèse. J'ai énormément appris à tes côtés et je te suis infiniment reconnaissante pour tout le temps que tu m'as accordé. Un grand merci à Adeline Courseaux pour ton soutien, ta bonne humeur et ta bienveillance, ainsi qu'à Dr Xavier Roussel pour tous les moments conviviaux lors des ateliers de médiation scientifique et des fêtes de la science ! À ce propos, je remercie également Dr Corentin Spriet, l'initiateur de ces projets formidables (Blob, bioplastique, etc.). Grace à vous, je ne suis pas prête d'oublier ces années formidables et ces nombreux resto au Coup d'Food !

Je remercie également ma fidèle binôme de paillasse, Dr Rayan Osman, pour tous ces moments partagés, nos longues soirées de purification ou sur les lignes SWING et DISCO du Synchrotron. Merci pour ton soutien indéfectible et tes encouragements depuis le master jusqu'à la rédaction de cette thèse.

Je remercie l'ensemble de l'équipe PlastoPol. Merci à Dr Christophe d'Hulst pour m'avoir accueillie dans son équipe pendant plus de trois ans. Merci également à Dr Fabrice Wattebled pour ses conseils lors de ma préparation au concours de l'école doctorale, ainsi que pour avoir accepté de participer à mon comité de suivi de thèse. Je tiens également à remercier Camille Loquet et Océane George pour leur soutien, leur bonne humeur, nos pauses thé, et nos déjeuners devant les films de Noël.

Je remercie également toute l'équipe administrative de l'UGSF ainsi que Dr Yann Guérardel pour son soutien dans tous mes projets (bourse L'Oréal, Xpérium). J'ai notamment beaucoup appris grâce à ma participation aux commissions départementales de biologie. Merci également pour l'attention et le soutien que tu portes à chacun des doctorants du laboratoire.

Mes remerciements s'adressent également aux membres du jury de thèse. Je remercie le Dr Tony Lefebvre d'avoir accepté de présider cette commission, ainsi que Dr Mirjam Czjzek et Dr Véronique Receveur Brechot pour avoir examiné mon manuscrit. Je remercie aussi Dr Arnaud Gruez et Dr Julie Bouckaert pour avoir accepté de participer en tant qu'examinateurs. Je vous adresse également mes sincères remerciements à tous, pour vos conseils avisés que ce soit au laboratoire, sur les lignes du synchrotron SOLEIL, lors de mon comité de suivi de thèse ou encore pendant les Workshops ou congrès.

Enfin mes remerciements s'étendent à l'ensemble de ma famille et belle-famille chez qui j'ai pu trouver tout le soutien nécessaire et les encouragements. Je tiens en premier lieu à remercier chaleureusement mes parents, en particulier ma mère qui a largement contribué à la réussite de mon cursus scolaire et mon père qui m'a inculqué la valeur du travail.

Merci à mes petites sœurs, Eve, Léa et Dafné, qui remplissent ma vie de bonheur. J'ai une immense chance de vous avoir dans ma vie. Je suis fière de chacune d'entre vous.

Merci à Victor. Ton soutien indéfectible, ta patience, et ton amour m'ont porté à travers toutes les étapes de ma vie depuis notre rencontre.

Merci à mes amis, Amélie D, Maé M, Jade M, Chloé M, Chloé H, Solène E, Virginie G, Amélie R, Zoé G, Eloise F, Caroline R, Océane G, Emeric L.

Je souhaite enfin rendre hommage à ma grand-mère, qui nous a quittés bien trop tôt. Elle a toujours été si fière de ses petites-filles, et je sais qu'elle aurait été profondément émue de me voir arriver jusqu'ici. Son amour et ses encouragements résonnent en moi encore aujourd'hui.



Merci à tous

Avant-propos

Avec une appétence pour la biologie, je me suis naturellement dirigée vers la licence Sciences de la vie à l'Université de Lille. Dès la deuxième année, j'ai choisi le parcours Biochimie pour me spécialiser dans l'étude de la fonction des molécules du vivant. J'ai également suivi l'option initiation à la recherche qui m'a permis d'approcher le monde de la recherche et qui m'a confortée dans le choix de faire une thèse de Biochimie. À la suite de ma licence, j'ai intégré le Master Chimie et sciences du vivant parcours sciences analytiques pour me spécialiser dans les techniques analytiques modernes employées pour la séparation et l'identification structurale des biomolécules. C'est au cours de ce master que j'ai pu découvrir notamment les approches de biologie structurale et biophysique enseignées par le Dr. Coralie Bompard. Mon attrait pour ces deux disciplines m'a poussée à effectuer mon stage de fin d'étude au sein de son équipe PlaStoPol. Cette équipe s'intéresse à l'élucidation des mécanismes du métabolisme de l'amidon et sa régulation. Mon stage portait sur la production d'une protéine impliquée dans l'initiation de la biosynthèse de l'amidon en vue de son étude structurale et qui constitue le projet de thèse proposé par le Dr. Coralie Bompard. À l'issue de ce stage, j'ai candidaté et obtenu une bourse de thèse auprès du concours de l'école doctorale de Biologie Santé de Lille en cofinancement avec la Région des Hauts-de-France. Ce projet de thèse s'inscrit donc dans la continuité de mes travaux de Master 2 et fait appel à de nombreuses disciplines de biologie moléculaire, biochimie, biologie structurale, biophysique et bio-informatique. Cette thèse multidisciplinaire fut un projet particulièrement complet d'un point de vue technique et captivant tout au long de mon doctorat.

Abréviations

Aa	Acide aminé
ADPG	ADP-glucose
AF	AlphaFold
AI	Auto-induction
AGPase	ADP glucose pyrophosphorylase
AMY	a amylase
At	Arabidopsis thaliana
BAM	βamylase
BE	Enzyme de branchement
BSA	Albumine sérum bovin
СВМ	Module de liaison aux sucres
CC	Coiled-coils
CD	Dichroïsme circulaire
DBE	Enzyme de débranchement
DP	Degré de polymérisation
DT	Domaine transmembranaire
DTT	Dithiothréitol
DPE2	Disproportionating enzyme 2
E.coli	Escherichia coli
ESV1	Early Starvation 1
F6P	Fructose-6-phosphate
FIP1	FTSH5 interacting protein 1
G6P	Glucose-6-phosphate
GBSS	Granule bound starch Synthase
GFP	Green fluorescent protein
GRAVY	Moyenne Générale d'hydropathie
GT	Glycosyl-transferase
GWD	α-glucan water dikinase
Hv	Hordeum vulgare
IEX	Chromatographie d'échange d'ions
ISA	Isoamylase
IMAC	Chromatographie d'affinité aux ions métalliques immobilisés
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny broth
LESV	Like Early Starvation 1
LSF2	Like starch excess four 2
MEX1	Maltose excess protein 1
MFP1	Mar binding filament like protein 1
MOS	Malto-oligosaccharides
Os	Oryza sativa
PAE	Predicted Aligned Error
PCR	Réaction en chaîne par polymérase
PHS1	α-glucan phosphorylase
Pi	Pyrophosphate
PII1	Protéine impliquée dans l'initiation 1

Predicted local distance difference test
Protein Targeting to Starch
Phosphoglucan water dikinase
RuBP carboxylase
Ribulose 1,5 bisphosphate
Domaine de dimérisation
Electrophorèse sur gel polyacrylamide en conditions dénaturantes
Chromatographie d'exclusion stérique
Starch excess 4
Rayonnement synchrotron
Solanum tuberosum
Amidon synthase
Translocation de l'arginine jumelle
2,2,2-Trifluoroéthanol
Type sauvage
Yellow fluorescent protein
3-phosphoglycérate

Table des matières

	Resume	3
	Remerciements	5
	Avant-propos	7
	Abreviations	8
	TABLE DES MATIERES	10
	TABLE DES FIGURES	13
	TABLE DES TABLEAUX	15
	Preambule	16
11	ITRODUCTION	18
	CHAPITRE 1. CARACTERISTIQUES GENERALES DE L'AMIDON	19
	1.1 Deux types d'amidon	19
	1.2 L'amidon est composé de deux polyglucanes	21
	1.3 Structure de l'amylose et de l'amylopectine	22
	1.4 Morphologie des grains d'amidon	24
	CHAPITRE 2. METABOLISME DE L'AMIDON	26
	2.1 Généralités	26
	2.2 Synthèse du précurseur de glucose à partir de la photosynthèse	28
	2.3 Les amidon synthases	30
	2.3.1 Les amidons synthases possèdent des domaines communs	30
	2.3.2 L'amidon Synthase liée aux grains	32
	2.3.3 Les amidons synthases solubles	34
	2.4 Branchement des chaînes d'amylopectine	35
	2.5 Débranchement des chaînes d'amylopectine	37
	2.6 LESV et ESV1 sont impliquées dans la formation et la protection des grains d'amidon	38
	2.7 Dégradation de l'amidon	44
	CHAPITRE 3. INITIATION DE LA BIOSYNTHESE DE L'AMIDON	46
	3.1 SS4 est une enzyme cruciale pour l'initiation des grains d'amidon	46
	3.1.1 SS4 est impliquée dans la taille, la forme et l'initiation de la biosynthèse des grains d'amidon	46
	3.1.2 SS4 n'est pas la seule enzyme capable d'initier	48
	3.1.3 SS4 génère des amorces pour les autres amidon synthases	49
	3.1.4 Rôles distincts des domaines de SS4 dans l'initiation	49
	3.2 Un complexe d'initiation impliquant des interactions protéine-protéine	51
	3.2.1 SS4 interagit avec plusieurs protéines dans le chloroplaste	51
	3.2.2 PTST2 apporte les MOS à SS4 et interagit avec PII1	54
	3.2.3 MFP1 localise le complexe aux thylakoïdes	55
	3.2.4 PII1 contrôle le nombre d'évènement d'initiation	58
	3.2.5 Résumé et connaissances actuelles sur l'initiation	60

OBJECTIFS	62
MATERIEL ET METHODES	64
CHAPITRE 1. MATERIELS ET SOLUTIONS UTILISEES	65
1.1 Souches bactériennes	65
1.2 Vecteurs	66
1.3 Solutions d'usage courant	66
1.4 Milieux de culture	67
CHAPITRE 2. PREPARATION DES PLASMIDES ET BACTERIES	68
2.1 Clonages	68
2.2 Préparation des bactéries chimio-compétentes	69
2.3 Transformations bactériennes	70
CHAPITRE 3. PREPARATION DES ECHANTILLONS PROTEIQUES	70
3.1 Induction des protéines	70
3.2 Lyses cellulaires	73
CHAPITRE 4. PURIFICATION DES ECHANTILLONS PROTEIQUES	74
4.1 Purification par chromatographie d'affinité sur ions métalliques immobilisés (IMAC)	74
4.2 Purification par chromatographie d'exclusion stérique (SEC)	74
4.3 Renaturation des corps d'inclusion	75
CHAPITRE 5. IDENTIFICATIONS DES PROTEINES	76
5.1 SDS-PAGE	
5.2 Western blot	76
5.3 Spectrométrie de masse	77
Chapitre 6. Etude de l'activite de SS4	77
CHAPITRE 7. ETUDE STRUCTURALE ET BIOPHYSIQUE	79
7.1 Dichroïsme circulaire	79
7.2 Etudes bio-informatiques	80
2.2.1 Analyse des séquences peptidiques	80
7.2.2 Prédiction des structures secondaires	80
7.2.3 Modélisation structurale assistée par intelligence artificielle	81
7.2.4 Alignements multiples des séquences	83
7.2.5 Prédiction coiled-coils	83
7.2.6 Prédiction des résidus conservés	83
RESULTATS	84
CHAPITRE 1. ANALYSE DE L'EXPRESSION DE PII1	86
1.1 Optimisation des conditions d'expression et de la solubilité de PII1	86
1.1.1 Tests d'expression	86
1.1.2 Expression de PII1-pET300	91
1.2 Realisation et expression des constructions tronquées de PII1	94
1.2.1 Clonage des constructions tronquées	94

1.2.2 Tests d'expression	96
1.2.3 Purification de PII1-H2H3	98
CHAPITRE 2. ANALYSE DE L'EXPRESSION DE SS4	.103
2.1 Clonage des constructions tronquées SS4	. 103
2.2 Tests d'expression des constructions tronquées de SS4	. 105
2.3 Purification par chromatographie d'exclusion stérique	. 107
CHAPITRE 3. FORMATION DU COMPLEXE PII1/SS4	.110
3.1 Co-purification des constructions tronquées de SS4 et PII1	. 110
3.1.1 Co-purification par affinité	110
3.1.2 Co purification par chromatographie d'exclusion stérique	112
3.2 Co-expression de PII1 et SS4	. 114
3.2.1 Purification du complexe par affinité	114
3.2.2 Purification du complexe par chromatographie d'exclusion stérique	116
CHAPITRE 4. ANALYSE STRUCTURALE	.127
4.1 Analyses bio-informatiques de la séquence de PII1	. 128
4.1.1 Analyse des propriétés physico-chimiques	128
4.1.2 Prédiction des éléments de structure secondaire	129
4.1.3 Prédiction de coiled-coils	130
4.1.4 Prédictions structurales de PII1	131
4.1.6 Analyse de l'alignement de séquence avec les orthologues	135
4.1.7 Analyse de la structure des orthologues avec AlphaFold3	137
4.1.8 Analyse structurale de PII1-H2H3 par dichroïsme circulaire par rayonnement synchrotron	138
4.2 Prédictions structurales de SS4, SS5 et des complexes PII1/SS4 et PII1/SS5	. 141
4.2.1 Analyse de la structure de SS4 dans une forme monomérique	141
4.2.2 Alignement de la région de dimérisation entre SS4 et SS5	144
4.2.3 Analyse de la structure de SS4, SS4 Δ 349 et SS5 sous forme dimérique	145
4.2.4 Prédiction structurale de l'interaction entre PII1 et SS4	148
4.3 Prédictions structurales de PTST2, PTST3, SS3 et MFP1	. 149
CHAPITRE 5. ETUDE ENZYMATIQUE DE PII1 ET SS4	.153
5.1 L'activité catalytique d'élongation de SS4 est fonctionnelle	. 153
5.2 PII1 inhiberait l'activité catalytique de SS4	. 156
DISCUSSION	. 159
PERSPECTIVES	.169
REFERENCES	.173
ANNEXES	.184

Table des figures

Figure 1. Les deux types de plaste	20
Figure 2. Chaînes de résidus de glucoses de l'amidon	22
Figure 3. Organisation de la structure du grain d'amidon.	24
Figure 4. Morphologies des grains d'amidon	25
Figure 5. Résumé synoptique du métabolisme de l'amidon.	28
Figure 6. Composition interne du chloroplaste	28
Figure 7. Structures prédites des amidon synthases	31
Figure 8. Structures prédites de GBSS et PTST1	34
Figure 9. Structures prédites des enzymes de branchement d'Arabidopsis thaliana	36
Figure 10. Structures prédites des enzymes de débranchement	37
Figure 11. Phénotypes des grains d'amidon de la lignée sauvage et des lignées mutantes LESV et ISA	40
Figure 12. Structure des protéines LESV et ESV1	43
Figure 13. Proposition du mécanisme d'action de LESV et ESV1	44
Figure 14. Phénotypes des grains et rosettes de la plante sauvage et du mutant ss4	48
Figure 15. Structure prédite de SS4	50
Figure 16. Phénotypes des grains d'amidon des lignées SS4 et ss4.	51
Figure 17. Phénotypes des grains d'amidon des lignées mutantes et doubles mutantes des protéir	nes
intervenant dans l'initiation	53
Figure 18. Structure prédite de PTST2 et PTST3	55
Figure 19. Structure prédite de MFP1 impliquée dans l'initiation	56
Figure 20. Structure prédite de PII1 impliquée dans l'initiation.	58
Figure 21. Interactions des protéines intervenant dans l'initiation.	60
Figure 22. Expression et solubilité de PII1	87
Figure 23. Optimisation de l'expression de PII1 en fonction de la souche et du vecteur d'expression	88
Figure 24. Optimisation de l'expression de PII1 par auto-induction.	90
Figure 25. Expression et caractérisation de PII1-pET300 produite dans BL21(DE3)	92
Figure 26. Optimisation de la solubilité de PII1 par l'ajout d'additifs dans le tampon de lyse	93
Figure 27. Analyse structurale de PII1 et conception des constructions tronquées	95
Figure 28. Expression des constructions tronquées de PII1.	97
Figure 29. Purification par chromatographie d'affinité de PII1-H2H3-pET300 sur colonne IMAC HisTrap Exce	əl 5
ml	99
Figure 30. Purification par chromatographie d'exclusion stérique de PII1-H2H3-pET300 sur colonne Super	dex
200 10/300 GL	00
Figure 31. Optimisation du choix du tampon d'élution pour la purification de PII1-H2H3 par chromatograp	hie
d'exclusion stérique sur Colonne Superdex 200 10/300 GL1	01

Figure 32. Analyse SDS-PAGE de la purification par chromatographie d'affinité de SS4 pDest17 produite dans
BL21(DE3)104
Figure 33. Description schématique des constructions tronquées de SS4105
Figure 34. Tests d'expression des constructions tronquées SS4 pETDuet106
Figure 35. Purification et caractérisation de SS4 Δ 349, SS4 Δ 403 et SS4 Δ 466
Figure 36. Purification des constructions tronquées SS4 pETDUET par chromatographie d'exclusion stérique
sur colonne Superdex 200 10/300 GL108
Figure 37. Expression des constructions tronquées PII1 et SS4 co-purifiées111
Figure 38. Co-purification de PII1-H2H3 et SS4∆349 (A,B,C) ou PII1-H2H3 et SS4∆403 (D,E) par
chromatographie d'exclusion stérique sur colonne Superdex 200 10/300 GL113
Figure 39. Purification du complexe PII1-SS4∆349 par affinité sur résine de Nickel
Figure 40. Purification du complexe PII1-SS4A349 par chromatographie d'exclusion stérique Superdex 200
10/300 GL
Figure 41. Prédiction in silico des éléments de structure secondaire à partir de la structure primaire de PII1
d'Arabidopsis thaliana avec (A) PSIPRED (McGuffin et al 2000) et (B) Gor4 (J Garnier et al 1996)
Figure 42. Prédiction de coiled-coils à partir de la structure primaire de PII1 d'Arabidopsis thaliana avec
MarCoil (Delorenzi and Speed, 2002)
Figure 43. Modèles de PII1 d'Arabidopsis thaliana calculés par (A) AlphaFold2 et (B,C,D) MassivFold132
Figure 44 . Modèles de (A) PII1 d'Arabidopsis thaliana et (B,C) PII1-H2H3 calculés par AlphaFold3
Figure 45. Analyse des régions et résidus conservées chez PII1 d'Arabidopsis et les orthologues
Figure 46. Superposition des modèles AlphFold3 de la structure hélices H2 et H3 de PII1 avec ceux des 28
orthologues calculés par AlphaFol3
Figure 47. Purification et analyse SR-CD de PII1-H2H3139
Figure 48. Influence de l'augmentation de la quantité de TFE sur PII1-H2H3 par SR-CD
Figure 49. Modèles de SS4 calculés par (A) AlphaFold2, (B) MassivFold et (C) AlphaFold3142
Figure 50. Superposition de la structure du domaine catalytique (GT1 et GT5) de SS4 résolue par
cristallographie (en vert) avec le domaine catalytique et la région de dimérisation (RD) de SS4 calculée par
AlphaFold3 (en bleu)
Figure 51. Alignement de structure entre SS4 et SS5144
Figure 52. Modèles AlphaFold3 en dimère de (A) SS4, (B) SS4∆349 et (C)SS5
Figure 53. Modèle de structure du complexe entre SS4 SS4∆349 (vert) et PII1-H2H3 (magenta) par MassivFold.
Figure 54. Modèles AlphaFold3 de PTST2 (A), PTST3 (B), MFP1 (C) et SS3 (D)
Figure 55. Caractérisation des activités d'élongation de SS4
Figure 56. Zymogramme révélant l'effet de PII1-H2-H3 sur l'activité d'élongation de SS4Δ349 et SS4Δ466.
Figure 57. Influence du rapport SS4Δ349/ PII1-H2H3 sur l'activité d'élongation de SS4
Figure 58. Proposition du mécanisme d'initiation de la biosynthèse du grain d'amidon chez la plante168

Table des tableaux

Tableau 1. Amorces nucléotidiques utilisées pour la construction des protéines tronquées de PII1. 69
Tableau 2. Récapitulatifs des conditions d'expression testées pour chaque construction protéique
Tableau 3. Compositions des milieux auto-inductibles utilisés pour l'optimisation de l'expression et de la
solubilité de PII1-pET300 dans BL21(DE3)
Tableau 4. Conditions d'expression et de purification utilisées pour l'expression PII1 entière118
Tableau 5. Conditions d'expression et de purification utilisées pour PII1-H2H3. 119
Tableau 6. Conditions d'expression et de purification utilisées pour PII1Δ71, PII1Δ101 et PII1-H1H2H3120
Tableau 7. Conditions d'expression et de purification utilisées pour la construction SS4Δ349. 121
Tableau 8. Conditions d'expression et de purification utilisées pour la construction SS4Δ403. 122
Tableau 9. Conditions d'expression et de purification utilisées pour la construction SS4Δ466. 123
Tableau 10. Conditions d'expression et de purification utilisées pour les co-purifications entre PII1-H2H3 et
les constructions tronquées de PII1 et SS4124
Tableau 11. Conditions d'expression et de purification utilisées pour les co-purifications entre les
constructions tronquées de PII1 et SS4125
Tableau 12. Conditions d'expression et de purification utilisées pour le double clonage PII1- SS4Δ349126
Tableau 13. Caractéristiques physico-chimiques et composition en acide aminés de PII1. 128
Tableau 14. Orthologues de PII1 d'Arabidopsis thaliana utilisés pour l'alignement de séquence multiple135

Préambule

Ce projet de thèse vise à étudier le rôle des protéines impliquées dans la biosynthèse de l'amidon à travers deux projets distincts. Le premier constitue la majeure partie de ma thèse et se concentre sur la protéine PII1, impliquée dans l'initiation de la biosynthèse de l'amidon. Le deuxième projet, auquel j'ai participé, porte sur les protéines LESV et ESV1, respectivement responsables de l'assemblage des molécules d'amylopectine dans le grain d'amidon et de sa protection contre une dégradation précoce.

Le sujet principal de ma thèse repose sur la compréhension du mécanisme d'initiation de la biosynthèse de l'amidon chez *Arabidopsis thaliana (At)*, en utilisant des approches biochimiques, biophysiques et structurales. L'objet de cette étude est la protéine PII1, protéine non enzymatique, dont le rôle dans l'initiation a été mis en évidence au laboratoire. PII1 interagit avec SS4, l'enzyme principale intervenant dans cette étape de la biosynthèse de l'amidon au sein d'un complexe multiprotéique d'initiation. L'étude du complexe PII1/SS4 et de l'influence de PII1 sur l'activité de SS4 ont également été réalisées dans ce travail.

A mon arrivée au laboratoire, la structure du domaine catalytique de SS4 était connue et la protéine entière avait pu être exprimée dans plusieurs travaux précédents. Certains travaux non publiés au laboratoire suggéraient une protéolyse entre le domaine Nterminal et le domaine catalytique C-terminal. Ces premiers essais d'expression de la protéine PII1 réalisés au laboratoire semblaient indiquer que, bien que difficile à produire, PII1 montrait des résultats encourageants quant à sa solubilité selon les conditions d'expression utilisées.

Malgré tout, L'optimisation des conditions d'expression pour obtenir les protéines PII1 et SS4 sous une forme soluble, nécessitant la réalisation de constructions tronquées, a représenté plus de la moitié de mon travail de thèse. Des approches de modélisation moléculaire ont également contribué à la définition des constructions tronquées et à l'analyse de la structure de PII1 et du domaine N-terminal de SS4.

Une fois ces constructions réalisées, la deuxième partie de ma thèse a été consacrée à leur purification et la caractérisation des protéines seules ainsi qu'à celle du complexe PII1/SS4. Ces échantillons protéiques ont également permis de tester l'activité enzymatique des constructions tronquées de SS4 et de démontrer que PII1 inhibe cette activité. Ces travaux ont permis de progresser dans la compréhension des mécanismes moléculaires au cours de l'initiation de la biosynthèse de l'amidon. Ces travaux ont fait l'objet d'un article récemment soumis dans lequel je suis première auteure (Annexe 1).

Le second projet portait sur l'étude des protéines LESV et ESV1 d'*Arabidopsis thaliana*. Ce travail constrituait le sujet de thèse de Rayan Osman (RO) au laboratoire, soutenue en juillet 2023. RO et moi avons beaucoup collaboré sur nos projets respectifs, notamment sur les collections de données sur plusieurs lignes de lumière du synchrotron SOLEIL et le traitement de ces données. Ces résultats ont fait l'objet d'une publication dans laquelle je suis deuxième auteure (Annexe 2).

En dehors de mes travaux de thèse, j'ai enrichi ma formation scientifique par le biais d'un stage en bio-informatique sur AlphaFold et d'un workshop en biologie structurale (RéNaFobis) ainsi que par un congrès de Biologie Structurale Intégrative à Marseille où j'ai pu exposer mes travaux de thèse sous la forme d'un poster.

Enfin, j'ai activement contribué à la vie de l'Université de Lille et du laboratoire, en participant aux projets de la fête de la science, en étant impliquée dans la médiation scientifique en tant qu'animatrice de deux stands à l'Xperium, vitrine scientifique de l'université (100h en trois ans) et ambassadrice des écoles. J'ai encadré deux stagiaires (M1, L3) et également participé à la vie institutionnelle en siégeant en tant que membre nommée au comité du département de biologie de la faculté de science et technologie de Lille.

INTRODUCTION

Chapitre 1. Caractéristiques générales de l'amidon

1.1 Deux types d'amidon

L'amidon est le principal glucide de stockage chez les végétaux supérieurs. Il s'accumule sous forme de grains, insolubles dans l'eau, synthétisés dans les plastes (Meyer, 1883). Les plastes sont des organites spécifiques des cellules végétales. On peut différencier deux types de plastes selon leur localisation et leur fonction *in planta*: les chloroplastes et les amyloplastes (**Figure 1**). Les chloroplastes, issus des tissus chlorophylliens comme les feuilles, permettent une mise à disposition immédiate de l'amidon. C'est ce qu'on appelle l'amidon transitoire. Les amyloplastes, quant à eux, sont présents dans les organes non photosynthétiques tels que les tubercules et les racines. Ils lls assurent le stockage de l'amidon de réserve, qui est accumulé en grande quantité et sera mobilisé après la germination ou en cas de gel (Fincher, 1989).C'est ce qu'on appelle l'amidon de réserve.

L'amidon transitoire est synthétisé durant la phase diurne à partir du carbone assimilé par la photosynthèse. La plante suit deux étapes majeures : d'abord l'initiation de l'amidon à partir des produits issus de la photosynthèse puis la biosynthèse proprement dite. La nuit, l'amidon est dégradé pour alimenter l'ensemble des tissus de la plante en saccharose et fournir le substrat nécessaire à la respiration des feuilles.

Ce chapitre traite des généralités sur l'amidon. Le chapitre 2 abordera les différentes sous-étapes de la biosynthèse et la dégradation de l'amidon, tandis que le chapitre 3 détaillera le sujet principal de ma thèse : l'étude de l'initiation.



Figure 1. Les deux types de plaste. L'amidon contenu dans les chloroplastes est appelé amidon transitoire tandis que l'amidon contenu dans les amyloplastes est appelé amidon de réserve.

L'amidon constitue la principale source de calories utilisées dans l'alimentation humaine. De nombreux produits fabriqués à partir de l'amidon de réserve sont commercialisés à des fins alimentaires et non alimentaires (Ball et al., 1996). La France représente actuellement 30 % de la production européenne d'amidon, ce dernier étant produit essentiellement à partir de tubercules de pommes de terre. Il est prédominant dans deux secteurs : l'industrie alimentaire (56 % en boulangerie, pâtisserie, boissons, charcuterie) et non alimentaire (44 % en papeterie, cartons ondulés, pharmacie) (USIPA, https://www.usipa.fr/chiffres-cles/). Par ailleurs, l'amidon joue un 2022 rôle particulièrement significatif dans la transition énergétique, où il est considéré comme une ressource renouvelable et biodégradable. En effet, la conversion de l'amidon en sucres fermentescibles permet la création de produits biosourcés tels que les bioplastiques et la production de bioéthanol (Smith, 2008).

La région des Hauts-de-France constitue la première région productrice de pommes de terre avec plus de 61% de la production française en 2017, ce qui confère à la région une retombée économique importante et de nombreux emplois (Agreste Mars 2019). L'un des problèmes rencontrés par les producteurs est la conservation des tubercules à froid, ce qui génère l'accumulation de sucres solubles et impacte la qualité organoleptique de l'aliment lors de la cuisson à haute température (Müller-Thurgau, 1882). La morphologie ou la structure moléculaire du grain influe sur les propriétés physico-chimiques applicables aux utilisations industrielles et peut donc influer sur la résistance au froid. Or la structure du grain est en partie établie durant le mécanisme de biosynthèse de l'amidon qui fait intervenir des protéines de l'initiation et de biosynthèse (Smith, 2008). Connaître le mécanisme général de la biosynthèse et de la dégradation de l'amidon permettra à terme d'améliorer les cultures à des fins alimentaires et non alimentaires.

1.2 L'amidon est composé de deux polyglucanes

L'amidon est un glucide complexe composé de polymères d'α-glucoses dont les deux principaux composants sont l'amylose et l'amylopectine. L'amylose est le composé minoritaire, représentant de 18% à 33% du poids sec. L'amylopectine est le composé majoritaire représentant 72 à 82% du grain (Banks et al., 1975). On sait aujourd'hui que l'amylose n'est pas indispensable à la formation du grain d'amidon. L'absence d'amylose dans la plante conduit à un phénotype caractérisé par des grains « cireux », appelés alors *waxy* (Nelson and Tsai, 1964; Weatherwax, 1922).

L'amylose et l'amylopectine sont constituées d'unités de glucose liées entre elles par des liaisons O-glycosidiques α -1,4 (**Figure 2**). Ces chaînes peuvent être ramifiées avec d'autres chaînes linéaires par des liaisons O-glycosidiques α -1,6 (**Figure 2**). Ces branchements représentent seulement 1 % de la structure de l'amylose contre 5 à 6 % pour l'amylopectine (Buléon et al., 1998; Hizukuri et al., 1981; Manners, 1989).



Figure 2. Chaînes de résidus de glucoses de l'amidon. Ces résidus sont liés par des liaisons α-1,4 et ramifiés par des liaisons α-1,6.

D'autres macromolécules rentrent dans la composition de l'amidon, c'est le cas des protéines et des lipides représentant respectivement 0,1 à 0,7% et 1,5% du poids sec du grain d'amidon (Biliaderis, 1979; Morrison et al., 1984).

1.3 Structure de l'amylose et de l'amylopectine

L'amylose est décrite comme ayant une structure principalement linéaire et peu branchée dans laquelle les chaînes de glucose s'agencent sous la forme d'hélices simples. On estime son poids moléculaire entre 10⁵ et 10⁶ Da (Buléon et al., 1998). La localisation et le rôle de l'amylose dans le grain ne sont pas bien compris. On suppose que les chaînes d'amylose s'intercalent entre les molécules d'amylopectine (Jane et al., 1994).

L'amylopectine est une structure plus complexe que l'amylose du fait de sa taille et de son organisation structurale. Ce polyglucane comporte entre 10^5 et 10^6 unités de glucose correspondant à un poids moléculaire d'environ 10^7 Da (Manners, 1989). Les résidus de glucose forment des chaînes linéaires d'une longueur comprise entre 6 et > 100 résidus (**Figure 3A**). On retrouve généralement trois groupes de chaînes qui se distinguent selon leur longueur (**Figure 3B**). La chaîne de type C, la plus longue, est celle qui se trouve au centre de la molécule et qui contient l'extrémité réductrice. Sur la chaîne de type C, vont se fixer d'autres chaînes de longueur intermédiaire (types B) par des liaisons α -1,6. Il existe trois sortes de chaîne de type B (B1, B2 et B3) qui varient selon le nombre de résidus de

glucose situés entre 10 et > 40 résidus. Les chaînes de type B3 sont les plus longues et sont ramifiées par les chaînes B2, plus courtes. Ensuite les chaînes B2 sont ramifiées par les chaînes B1 de longueur < B2. Enfin les chaînes B1 portent les chaînes de type A qui se situent à l'extrémité non réductrice de la chaîne d'amylopectine. Ces chaînes de type A sont les plus courtes, en moyenne 12 à 15 résidus, et ne contiennent pas de ramification (Bertoft, 1991; Hizukuri, 1986; Hizukuri et al., 1981).

L'ensemble des chaînes linéaires de taille variable forme une structure en grappe (Figure 3B). Les chaînes linéaires, agencées parallèlement, vont s'associer pour former des doubles hélices qui vont à leur tour s'agencer en réseau organisé (Figure 3C). On distingue deux types d'organisation A ou B des doubles hélices aboutissant à deux types de système cristallin spécifique à un tissu végétal (Figure 3D). Ainsi l'organisation selon le type A est retrouvée principalement dans l'amidon de réserve des céréales tandis que l'organisation selon le type B est retrouvée dans les tissus à forte teneur en eau, par exemple les feuilles ou les tubercules (O'Sullivan and Perez, 1999). Cet agencement en double hélice et structuration en grappe va créer une région cristalline qui, dans le grain, formera des couches (ou lamelles) cristallines concentriques (Figure 3E). D'autre part, les points de ramifications se retrouvent concentrés dans une région sans double hélice dite amorphe. C'est à ce niveau que logerait l'amylose.

Les lamelles cristallines alternent avec les lamelles amorphes avec une périodicité de 9 à 10 nm et confèrent la caractéristique semi-cristalline du grain d'amidon (Jenkins and Donald, 1995). Les lamelles prennent naissance au centre du grain d'amidon, autour d'un point central qu'on appelle le hile, puis elles s'accumulent radialement depuis le hile vers l'extérieur du grain. C'est aussi à ce niveau que le grain commence à s'organiser puis à cristalliser. On observe généralement le hile au microscope polarisé sous la forme caractéristique de la croix de malte car les chaînes d'amylopectine s'agencent perpendiculairement à la surface (Jane et al., 1994). Le hile est observable principalement dans les grains issus des amyloplastes ou céréales. Les grains d'amidon issus des feuilles d'*Arabidopsis thaliana* par exemple, sont plats et ne possèdent donc pas de hile caractéristique.



Figure 3. Organisation de la structure du grain d'amidon. (A) L'amylopectine est composée d'unités de glucose liées par des liaisons α-1,4 et ramifiées par des liaisons α-1,6. (*B*) L'amylopectine contient trois types de chaînes de glucose de longueur variable formant une structure en grappe. (*C*) Deux chaînes de glucose agencées parallèlement s'associent en doubles hélices. (*D*) Les doubles hélices d'amylopectine sont compactées en cluster et arrangées en type A ou B selon l'interaction des molécules d'eau avec l'amylopectine. (*E*) Les clusters de doubles hélices sont compactés au même endroit, formant une lamelle cristalline de 6 nm. Les lamelles cristallines alternent avec les lamelles amorphes de 3 nm constituées de l'ensemble des points de ramification. (*F*) L'alternance de ces lamelles cristallines et amorphes confère le caractère semi-cristallin au grain d'amidon et l'apparition d'anneaux de croissance. Figure adaptée de (Zeeman et al., 2010).

1.4 Morphologie des grains d'amidon

La taille des grains d'amidon peut varier de 1 μ m à plus de 100 μ m selon les espèces et les organes, avec des formes très variées (**Figure 4**) (Czaja et al., 1969). On retrouve par exemple de gros grains ronds de plus de 100 μ m dans les tubercules de pomme de terres (moyenne de 40 μ m) tandis que dans les feuilles d'*Arabidopsis thaliana (At)*, les grains d'amidon sont beaucoup plus petits, généralement 2 μ m de diamètre et de forme lenticulaire (Buléon et al., 1998; Zeeman et al., 2002). Certains grains sont de forme anguleuse ou sphérique, comme l'amidon de l'albumen de riz (*Oryza sativa*) ou de maïs (*Zea mays*). Chez le blé (*Tricum aestivum*), on observe également une distribution bimodale de grains avec la synthèse de gros granules de type A qui apparait 4 jours après l'anthèse et d'autres plus petits de type B, qui font leur apparitions 10 à 15 jours après l'anthèse (Jane et al., 1994).



Figure 4. Morphologies des grains d'amidon. Les grains d'amidon sont visualisés par microscope électronique à balayage. Figure adaptée de (Khalid et al., 2017). amidotheque.cermav.cnrs.fr

La morphologie du grain d'amidon pourrait être définie au départ de la croissance de la plante ou influencée par l'environnement dans lequel il est synthétisé. L'espace disponible dans le stroma pourrait influencer la taille des grains (Crumpton-Taylor et al., 2012; Crumpton-Taylor et al., 2013). Par exemple, dans les feuilles d'*Arabidopsis*, la synthèse des grains démarre à proximité des thylakoïdes. A cet endroit, la place disponible pour la synthèse d'un grain est restreinte et ne permet que la formation de petits grains (Bürgy et al., 2021; Gámez-Arjona et al., 2014). En revanche, dans les tubercules de pommes de terre, la place disponible pour la synthèse des grains d'amidon au sein des amyloplastes est plus importante par rapport à celle dans les feuilles d'*Arabidopsis*, ce qui permet la formation de grains plus gros. Récemment d'autres facteurs influençant la taille ou la forme du grain ont été identifiés. Il s'agit de protéines issues de la biosynthèse qui, lorsqu'elles ne sont pas exprimées dans la plante, altèrent le phénotype du grain (Gámez-Arjona et al., 2018). Ces changements impactent également les propriétés physico-chimiques du grain qui peuvent être exploitées pour les applications industrielles de l'amidon (Apriyanto et al., 2022; Jobling, 2004).

Chapitre 2. Métabolisme de l'amidon

2.1 Généralités

Le métabolisme de l'amidon comprend une étape de biosynthèse et de dégradation des grains d'amidon. La biosynthèse de l'amylopectine comprend en elle-même cinq étapes :

- l'initiation durant laquelle sont synthétisées les amorces qui vont servir de substrat pour les étapes suivantes ;

- l'allongement des amorces pour former des chaînes linéaires de glucanes ;

- la formation des points de branchement ;

- la réorganisation des points de branchement ;

Ces étapes se déroulent pendant le jour et aboutissent à la formation des grains d'amidon.

La nuit, l'amidon est dégradé par des étapes de phosphorylation/déphosphorylation des chaînes puis par le débranchement et l'hydrolyse des chaînes de l'amylopectine.

Toutes ces étapes nécessitent l'action de multiples enzymes et protéines qui agissent de façon concertée. Chaque protéine joue un rôle bien spécifique dont leur fonction dans la plante a été révélée en grande partie par la génétique inverse. Ces plantes mutantes voient l'expression d'un gène en particulier (codant pour une protéine spécifique), éteinte ou surexprimée afin d'identifier des différences phénotypiques entre la plante mutante et la plante sauvage. Ces différences s'observent notamment sur l'aspect physique général de la plante et sa croissance ou plus localement dans le chloroplaste en regardant l'aspect des grains d'amidon et leur composition en polyglucanes. De plus, le métabolisme décrit est principalement celui qui se produit dans les feuilles d'*Arabidopsis*, le modèle d'étude très utilisé en laboratoire. En revanche, la biosynthèse ainsi que la dégradation peuvent varier selon l'espèce végétale étudiée ou des tissus.

Le rôle des enzymes de la biosynthèse sera détaillé dans ce chapitre mais un aperçu général des enzymes impliquées dans la synthèse et la dégradation de l'amidon est résumé dans la **Figure 5**.

26



Figure 5. Résumé synoptique du métabolisme de l'amidon. (A) Capture de l'énergie lumineuse : Le jour, les photons sont captés par les pigments photosynthétiques des thylakoïdes. Cette énergie est convertie en NADPH et ATP, utilisés dans le cycle de Calvin pour fixer le CO2 sur le ribulose 1,5-bisphosphate (RuBP) via l'enzyme RuBisCo, produisant du triose-phosphate. Ce dernier est exporté vers le cytoplasme pour être transformé en saccharose. (*B*) Initiation de la biosynthèse de l'amidon : L'ADP-glucose, généré à partir du fructose-6-phosphate, un intermédiaire du cycle de Calvin, est utilisé par les amidon-synthases pour transférer les unités de glucose sur des amorces malto-oligosaccharidiques (MOS). La SS4 et plusieurs protéines non enzymatiques (PTST2, PTST3, MFP1, PII1) jouent un rôle crucial dans ce processus. (*C*) Allongement des chaînes d'amidon : La GBSS allonge les chaînes de MOS pour former l'amylose, tandis que les amidon synthases et les enzymes de branchement (BEs) et de débranchement (DBE) structurent l'amylopectine. LESV et ESV1 aident à organiser et protéger les grains d'amidon nouvellement synthétisés. (*D*) Dégradation nocturne de l'amidon : La nuit, la phosphorylation et déphosphorylation par GWD, PWD et SEX4 rendent l'amidon accessible aux DBE, β et α -amylases (BAM et AMY), qui le dégradent en MOS, maltose et maltotriose. Ces produits peuvent être réutilisés pour la synthèse d'amidon ou convertis en saccharose dans le cytosol.

2.2 Synthèse du précurseur de glucose à partir de la photosynthèse

La biosynthèse des grains d'amidon démarre avec la photosynthèse (**Figure 5A**). La photosynthèse est un processus biologique qui a lieu dans les chloroplastes et permet de convertir l'énergie lumineuse en énergie chimique à partir des tissus photosynthétiques des végétaux chlorophylliens (**Figure 6**). Le carbone inorganique présent dans l'atmosphère est converti en carbone organique qui sera métabolisé par la plante pour sa croissance.

Elle comprend deux phases : une phase claire également appelée phase photochimique qui se déroule dans la membrane des thylakoïdes et une phase obscure ou non photochimique, qui n'a pas besoin de lumière et se déroule dans le stroma du chloroplaste à partir des produits issus de la phase claire.



Figure 6. Composition interne du chloroplaste.

La lumière émet des rayonnements $h\nu$ qui seront captés par des pigments (chlorophylles et caroténoïdes) contenus dans les membranes des thylakoïdes. Cette étape permet de convertir l'énergie lumineuse reçue en ATP et NADPH, H+. Puis ces deux composés seront utilisés dans le stroma du chloroplaste en vue de transformer le CO₂ atmosphérique en glucose selon l'équation suivante :

$6CO_2 + 6H_2O \rightarrow C_6H_{12}O_6 + 6O_2$

Cette étape fait appel à un ensemble d'enzymes au cours du cycle métabolique appelé le cycle de Calvin-Benson-Bassham (Calvin and Benson, 1948). Le CO2 est absorbé par les stomates des tissus foliaires et incorporé dans le cycle de Calvin en se liant à une molécule acceptrice, le ribulose 1,5 bisphosphate (RuBP) par l'action de la RuBP carboxylase (RuBisCo). Il est réduit en triose-phosphates qui sont exportés vers le cytoplasme pour être métabolisés en saccharose redistribué dans tous les tissus comme source d'énergie et de carbone. L'ADP-glucose (ADPG), le donneur de glucose est généré directement par la conversion du fructose-6-phosphate (F6P), un intermédiaire du cycle de Calvin, en glucose-6-phosphate (G6P) par la phosphoglucose isomérase (PGI) puis en glucose-1-phosphate (G1P) par la phosphoglucomutase (PGM1). L'ADPG pyrophosphorylase (AGPAse) catalyse ensuite la réaction du G1P et d'ATP en ADPG et pyrophosphate (Pi). Ces étapes sont réversibles. L'activité de l'AGPase est régulée par deux effecteurs allostériques qui jouent un rôle clé dans la biosynthèse de l'amidon. L'activité est augmentée en présence de 3-PGA (un intermédiaire du cycle de Calvin) et inhibée par la présence de Pi. La synthèse de l'ADPG peut également se faire à partir du saccharose synthétisé dans le plaste et exporté dans le cytosol.

Dans l'albumen des graines de céréales, l'ADPG n'est pas plastidial mais exclusivement cytosolique et transporté dans l'amyloplaste pour la synthèse de l'amidon de réserve (Kirchberger et al., 2007).

L'ADPG sert de substrat aux différentes enzymes intervenant dans l'initiation de la biosynthèse de l'amylose et de l'amylopectine (Leloir et al., 1961; Recondo and Leloir, 1961). Le glucose est transféré sur les amorces de chaînes glycaniques linéaires préexistantes par les amidon-synthases (SSs). Dans les mutants d'*At*, l'absence d'AGPase, de PGI ou de PGM1, bloque la synthèse de l'ADPG qui ne peut plus être utilisé par les SSs pour la synthèse de nouveaux grains (Bahaji et al., 2011). De plus, l'AGPase permet de réguler le métabolisme de l'amidon sous stress osmotique ou lumineux en contrôlant le flux de carbone (Ballicora et al., 2000; Hendriks et al., 2003). Cependant, le rôle de l'AGPase dans l'ajustement à court terme (amidon transitoire) et long terme (amidon de réserve) du taux de synthèse de l'amidon n'est pas clair.

2.3 Les amidon synthases

2.3.1 Les amidons synthases possèdent des domaines communs

Chez les végétaux supérieurs, les SSs catalysent le transfert d'unité de glucose à partir de l'ADPG, à l'extrémité non réductrice d'une amorce malto-oligosacharidique (MOS) linéaire ou ramifiée pour former l'amylose et l'amylopectine (**Figure 5C**). Il existe six classes de SSs que l'on peut diviser en deux groupes :

- Granule Bound Starch Synthase (GBSS), liée au grain d'amidon, responsable de la biosynthèse de l'amylose (Shure et al 1983);
- Soluble Starch Synthases (SS1, SS2, SS3, SS4, SS5) responsables de la synthèse de l'amylopectine. Chez certaines espèces comme le maïs, on retrouve plusieurs isoformes désignées généralement par « a « ou « b ». Chez la pomme de terre on retrouve également l'isoforme SS6 (Helle et al., 2018; Yu et al., 2012). SS5 contrairement aux autres ne possède pas d'activité enzymatique.

Les amidon synthases contiennent une région catalytique commune très conservée et située à l'extrémité C-terminale (**Figure 7**).



Figure 7. Structures prédites des amidon synthases. (A) structure des domaines prédits des amidon synthases d'Arabidopsis thaliana (At). En bleu : les peptides de transit au chloroplaste prédits. En vert : les domaines coiled-coils prédits (C). En Jaune : les domaines de liaisons aux sucres (CBM). En noir : les domaines de glycosyl transferase 5 (GT5). En rouge : les domaines glycosyl transferase 1 (GT1). Figure adaptée de Pfister & Zeeman, 2016. (**B**) structure du domaine catalytique d'AtSS4 résolue par cristallographie par Nielsen et al 2018. La protéine a été cristallisée avec de l'ADP, du maltose et de l'acarbose (représentation en sticks verts) situés au niveau du site actif entre les deux domaines GT. En jaune : le motif KXGGL conservé chez les amidon-synthases responsables de l'interaction avec l'ADPG. La figure a été réalisée avec PyMOL avec la structure AtSS4 (Code 6GNE) déposée dans la banque de données protéiques (PDB).

Les SSs contiennent deux sous-domaines d'activité glycosyltransférase (GT1 et GT5) et de type Rossmann (CAZY). Le repliement de Rossman est constitué d'un feuillet β de 6 brins parallèles liés à deux paires d'hélices a dans l'ordre β – α – β . Ces deux sous-domaines sont reliés par une région charnière. Le site actif se trouve entre ces deux sous-domaines. Il a été proposé que GT1 se lie du substrat donneur (ADPG) et GT5 se lie au substrat accepteur (MOS)(Qasba et al., 2005). SS5, est la seule SS qui ne contient pas de GT1. Elle est donc considérée comme non catalytique (Abt et al., 2020). Les glycosyltransférases catalysent le transfert des groupes glycosylés de l'ADPG vers un accepteur nucléophile avec rétention de configuration. La structure de ce domaine catalytique est représentée **Figure 7B** à partir de la structure résolue du domaine catalytique de la protéine *At*SS4 (Nielsen et al., 2018).

D'autres structures du domaine catalytique ont été résolues dans l'orge (*Hordeum vulgare*) pour *Hv*SS1 (Cuesta-Seijo et al., 2016) et dans le riz pour *Os*GBSS1 (Momma and Fujimoto, 2012). Ces protéines possèdent toutes un motif K-X-G-G-L conservé qui sert de site de liaison à l'ADPG (Furukawa et al., 1990). Pour reproduire cette interaction et les stabiliser, les protéines ont été cristallisées avec un substrat ADPG et du maltose (pour

*At*SS4) ou maltotriose (pour *Hv*SS1). *At*SS4 a également été cristallisée avec de l'acarbose, un inhibiteur de la glucosidase utilisé comme mimétique d'accepteur de glucose (**Figure 7B**). La liaison à l'ADP se fait par des liaisons hydrogènes (**Figure 7B**). Les auteurs ont mis en évidence deux résidus conservés qui pourraient servir de liaison avec les sucres (le tryptophane 688 et la phénylalanine 754).

Les domaines N-terminaux des SSs ne sont pas conservés et diffèrent en longueur et en composition. Les SSs possèdent également un peptide de transit prédit de longueur variable permettant l'adressage au chloroplaste. Ce domaine N-terminal contient principalement des hélices agencées en coiled-coils (CC) (**Figure 7**) et des CBM. On observe chez SS4, des domaines CC situés entre 204 et 466 acides aminés. Ces CC sont connus pour être impliqués dans les interactions protéines-protéines. Dans les mécanismes de biosynthèse, les SSs interagissent avec une voire plusieurs protéines pour former des hétérocomplexes ou homocomplexes (Burkhard et al., 2001). La SS3 possède trois domaines de liaison aux sucres (CBM53) prédits et qui servent à la liaison de la protéine avec les chaînes d'amylopectine ou les MOS. Enfin, SS4 et SS5 ont été récemment décrites avec une région qui sert à sa dimérisation, appelée région de dimérisation RD et située entre le domaine N-terminal et C-terminal (**Figure 14**) (Abt et al., 2020; Raynaud et al., 2016).

2.3.2 L'amidon synthase liée aux grains

La GBSS est la protéine majoritaire contenue dans le grain d'amidon où elle représente 85% du total des protéines associées aux grains (Echt and Schwartz, 1981) Elle allonge les chaînes de glucanes en rajoutant des unités de glucoses à l'extrémité non réductrice par des liaisons en α -1,4 à partir de l'ADPG (Leloir et al., 1961; Rongine De Fekete et al., 1960). Le mécanisme de ramification des chaînes linéaires de glucane de l'amylose n'est pas encore clair mais on suggère qu'il pourrait être effectué par les Enzymes de Branchement (BE) non pas situées dans le stroma mais plutôt liées aux grains (Tetlow, 2004). De plus, il a été montré que, *in vitro*, la GBSS peut utiliser l'amylopectine et les MOS. L'amylopectine sert non seulement de substrat à la GBSS, mais agit également comme un

effecteur de cette enzyme, accélérant ainsi l'élongation des MOS (Denyer et al., 1999; Seung et al., 2015).

Toutefois, la biosynthèse de l'amylose n'est pas indispensable à la formation du grain d'amidon. En effet, chez le maïs, la plante mutante qui ne produit pas la GBSS responsable de la synthèse de l'amylopectine montre un phénotype normal du grain d'amidon mais sans amylose (Sprague et al., 1943). La formation du grain est strictement liée à la biosynthèse de l'amylopectine.

Chez *Arabidopsis thaliana*, une autre protéine est requise pour la synthèse de l'amylose, il s'agit de la Protein Targeting to Starch (PTST1). PTST1 est une protéine plastidiale non enzymatique qui interagit avec la GBSS (Seung et al., 2015). Les protéines PTST appartiennent à la famille des « protéines ciblant l'amidon » et contiennent trois membres : PTST1, PTST2et PTST3. Ces deux dernières servent à cibler l'amylopectine pendant l'initiation et seront décrites dans le chapitre 3.

PTST1 interagit physiquement avec la GBSS *via* une région CC présente dans sa structure et dans celle de la GBSS (**Figure 8**) pour former un complexe protéine-protéine stable. Le prédicteur de CC COILS a prédit une petite région située en C-terminal de la GBSS pouvant être impliquée dans l'interaction avec PTST1 (Hélice orange sur la **Figure 8**). Sur la structure prédite de la GBSS par AlphaFold (AF), la région se traduit par une petite hélice de 14 acides aminés (Seung et al., 2015) (**Figure 8**). PTST1 contient en plus un CBM48 qui permet la localisation correcte de la GBSS dans les grains d'amidon et la plante produit moins d'amylose (Seung et al., 2015). Ces résultats démontrent qu'au cours de la biosynthèse, les SSs interagissent avec d'autres protéines de la biosynthèse de l'amidon, même sans activité catalytique.



Figure 8. **Structures prédites de GBSS et PTST1.** (*A*) Les domaines prédits de GBSS (haut) et PTST1 (bas) avec les coiled-coils (en orange) prédits par le programme COILS. (*B*) Modèles AlphaFold2 de GBSS (gauche) et PTST1 (droite). L'hélice de GBSS impliquée dans les CC est colorée en orange. Le CBM de PTST1 impliqué dans la liaison avec les sucres est coloré en rouge. TP = peptide de transit, CC = coiled-coils, CBM = module de liaison aux sucres. Adaptée de (Seung et al., 2015)

2.3.3 Les amidons synthases solubles

Les études montrent que chaque SS a une affinité spécifique pour leur substrat en fonction de son degré de polymérisation (DP). L'absence de l'une d'elles dans la plante altère significativement la structure finale de l'amylopectine (Delvallé et al., 2005; Zhang et al., 2008, 2005). Ainsi SS1 allonge préférentiellement les chaînes courtes de DP 6 ou 7 (Delvallé et al., 2005). SS2 catalyse l'élongation des chaînes moyennes de DP 13 à 20 (chaînes A et B1 **Figure 3B**) (Zhang et al., 2008) permettant la formation des doubles hélices. Chez les céréales comme le maïs, on a identifié deux isoformes (SS2a et SS2b). SS2a permet d'allonger les chaînes de 6 à 10 résidus en une chaîne de longueur intermédiaire de 12 à 25 résidus. Le rôle de SS2b n'est pas connu (Harn et al., 1998; Zhang et al., 2005). SS3 participerait à la synthèse de longues chaînes B plus longues entre DP 25 et DP 35 et permettrait également d'assurer la synthèse d'amidon en absence de SS4 (Szydlowski et al., 2009).

SS4, n'a pas d'influence sur la structure de l'amylopectine (Roldán et al., 2007; Szydlowski et al., 2009). La protéine a été décrite comme étant un facteur important pour contrôler le nombre et la forme des grains d'amidon dans les chloroplastes (Roldan et al., 2007). Elle agit en amont des autres SSs au niveau de l'étape d'initiation des grains d'amidon plutôt que dans l'allongement des chaînes de glucanes (Malinova et al., 2017; Roldán et al., 2007). Elle n'agit pas seule mais semble impliquée dans un complexe multiprotéique favorisé par la présence de CC (Liu et al., 2009; Seung et al., 2017; Tetlow et al., 2008; Vandromme et al., 2019). Enfin, SS5 a été récemment caractérisée comme étant non canonique contrairement aux autres isoformes. Elle n'interagirait pas directement avec les glucanes mais est impliquée dans l'initiation (Abt et al., 2020; Abt and Zeeman, 2020).

Les protéines SS4, SS3 et SS5 impliquées dans les mécanismes de l'initiation seront détaillées dans le chapitre 3.

2.4 Branchement des chaînes d'amylopectine

La deuxième classe d'enzyme qui intervient dans la biosynthèse de l'amidon concerne les Enzymes de Branchement (BEs pour Branching Enzymes). Les BEs appartiennent à la famille des α -amylases (CAZY). Ces enzymes hydrolysent la liaison α -1,4 d'un glucane et transfèrent la chaîne glycanique libérée vers une autre chaîne linéaire par une liaison α -1,6 (Mouille et al., 1996; Myers et al., 2000) (**Figure 5C**). De ce fait, les BEs génèrent des substrats supplémentaires pour les SSs. Le placement des points de branchement est crucial pour la structure de l'amylopectine mais le mécanisme de distribution reste incompris. Les BEs possèdent une région N-terminale contenant un CBM48, une région centrale composée d'un domaine α -amylase catalytique (AMY) et un domaine C-terminal typique de la famille des α -amylases (**Figure 9**).



Figure 9. Structures prédites des enzymes de branchement d'Arabidopsis thaliana. En bleu : les peptides de transit au chloroplaste. En Jaune : les domaines de liaisons aux sucres (CBM48). En noir : les domaines α -amylases catalytiques (AMY). En Bleu : Les domaines C-terminaux tout β conservés dans la famille des α -amylases. Figure adaptée de (Pfister and Zeeman, 2016).

Il existe deux classes d'enzyme de branchement, BE1 et BE2 (qui existe en deux isoformes BE2a et BE2b). Les BE1 transfèrent préférentiellement des chaînes plus longues que les BE2. Deux isoformes BE2 (BE2a, BE2b) sont présentes dans l'albumen de maïs (Boyer and Preiss, 1978; Fisher et al., 1993; Gao et al., 1997). L'absence de BE2b conduit à une réduction des chaînes courtes d'amylopectine et une augmentation de chaînes plus longues (Boyer and Preiss, 1978; Fisher et al., 1993). Chez *Arabidopsis*, on retrouve également deux isoformes BE de type II (BE2.2 et BE2.1) possédant une fonction redondante (Dumez et al., 2006). La perte simultanée de BE2.2 et BE2.1 chez *Arabidopsis* entraîne une perte totale de l'activité BE et les plantes doubles mutantes ne produisent plus d'amidon. A la place, une grande quantité de maltose s'accumule dans le chloroplaste (Dumez et al., 2006). Une troisième classe putative de BE (BE1) a été identifiée par homologie de séquence chez *Arabidopsis* (30% avec BE1 et BE2) mais cette isoforme n'intervient pas dans le métabolisme de l'amidon (Dumez et al., 2006).

Bien que les mécanismes moléculaires ne soient totalement compris, les BEs et les SSs peuvent former des complexes transitoires impliqués dans la voie de biosynthèse (Grimaud et al., 2008; Tetlow, 2004). D'autres complexes ont été identifiés dans les amyloplastes des céréales révélant des interactions fonctionnelles entre SS1 et SS2 et les enzymes de branchement BE2 (Ahmed et al., 2015; Crofts et al., 2015; Liu et al., 2009; Tetlow et al., 2008). L'activité de chaque enzyme prise indépendamment ne permet pas d'expliquer en totalité les mécanismes de biosynthèse de l'amidon. L'association de
plusieurs enzymes en complexe transitoire permet à la fois la colocalisation et l'action concertée (et certainement une régulation) de plusieurs activités enzymatiques qui pourrait expliquer la régulation spatio-temporelle et l'efficacité de ce mécanisme complexe.

2.5 Débranchement des chaînes d'amylopectine

Les enzymes de débranchement (DBE pour DeBranching Enzymes) contiennent deux classes d'enzymes : les limite-dextrinase (LDA) et les isoamylases qui présentent trois isoformes (ISA1, ISA2, ISA3) (Hussain et al., 2003)(Hussain et al 2003). Les DBE appartiennent à la famille des glycosyl hydrolases (GH 13) (CAZY). Elles hydrolysent les liaisons a- 1,6 de l'amidon et éliminent des points de branchements surnuméraires, ce qui permet l'homogénéisation de leur répartition au sein des molécules d'amylopectine (**Figure 5C**). Cette étape aboutit à une disposition alternant les points de branchement et les doubles hélices de façon régulière dans les chaînes d'amylopectine. Cette organisation permet la cristallisation des grains (Delatte et al., 2005; Myers et al., 2000).

Les glycosyl hydrolases partagent un domaine central α-amylase (AMY) et un CBM48 (**Figure 10**). Ces protéines sont adressées aux chloroplastes *via* un peptide de transit.



Figure 10. Structures prédites des enzymes de débranchement. Les enzymes de débranchement contiennent les isoamylases ISA et limite dextrinase LDA. En bleu : peptide de transit aux chloroplastes. En jaune : domaine de liaison aux sucres (CBM 48). En noir : domaine catalytique de glycosyl-hydrolase (AMY). En rouge : domaine PFam (DUF) de fonction inconnue. At : Arabidopsis thaliana. Figure adaptée de (Pfister and Zeeman, 2016).

ISA1 et ISA2 forment un complexe hétéromultimérique (Delatte et al., 2005; Wattebled et al., 2005). Les lignées dépourvues en ISA1 ou ISA2 montrent une désorganisation de l'amylopectine et une accumulation de polysaccharides solubles branchés de façon aléatoire appelés phytoglycogènes du fait de leur organisation proche de celle du glycogène, source de glucide des eucaryotes (James et al., 1995; Nakamura et al., 1996; Pan and Nelson, 1984).

Le phytoglycogène, possède des chaînes plus courtes et plus ramifiées et serait donc la forme soluble de l'amylopectine, avant l'action de ISA1 et ISA2. La présence de points de branchements au sein de la molécule empêche les molécules d'amylopectine de s'associer en double hélices et de s'organiser pour former un grain semi-cristallin. Cela démontre le rôle important des DBEs dans la formation des grains semi-cristallins (Streb et al., 2009). Deux autres protéines non enzymatiques, ESV1 et LESV, ont récemment été découvertes et jouent également un rôle dans la transition de phase de l'amylopectine, de sa forme soluble vers sa forme solide dans le grain (LESV) et dans le maintien de son intégrité au cours de la phase diurne (ESV1) (Feike et al., 2016; Liu et al., 2023; Osman et al., 2024).

2.6 LESV et ESV1 sont impliquées dans la formation et la protection des grains d'amidon

Early Starvation 1 (ESV1) et Like Early Starvation 1 (LESV) ont été découvertes récemment chez *Arabidopsis* (Feike et al., 2016; Liu et al., 2023; Malinova et al., 2018). Ces protéines, présentes dans le stroma du chloroplaste et liées aux grains d'amidon, jouent des rôles distincts : LESV intervient dans la transition de phase de l'amylopectine, tandis qu'ESV1 protège les grains nouvellement synthétisés de la dégradation précoce par les β/α -amylases (Liu et al., 2023; Osman et al., 2024).

Le renouvellement de l'amidon se produit la nuit, l'amidon est converti en saccharose selon un rythme presque linéaire, de sorte qu'environ 95% de l'amidon est consommé à l'aube. Ce rythme s'adapte à la durée du jour, optimisant le flux de carbone

pour la croissance de la plante. Quand la durée du jour diminue, les réserves disponibles en amidon sont régulées de façon à ne pas être épuisées avant la fin de la nuit (Gibon et al., 2004).

Le mutant *esv1* présente un phénotype avec une altération du renouvellement de l'amidon. Les réserves d'amidon sont épuisées 2h avant l'aube, entraînant une famine nocturne précoce et les grains présentent également une dégradation précoce pendant la phase diurne (Feike et al., 2016). *esv1* présente aussi une altération de la structure des grains d'amidon mais pas de la structure de l'amylose et l'amylopectine. La surexpression de ESV1 dans la plante entraine une accumulation d'amidon et la présence de grains d'amidon beaucoup plus gros que dans les plantes sauvages. L'ensemble de ces observations ont conduit à proposer qu'ESV1 joue un rôle dans la protection des grains et la prévention de leur dégradation précoce (Feike et al., 2016; Liu et al., 2023; Osman et al., 2024).

Nous avons vu dans la section 2.5 que l'absence de ISA1 induit une accumulation de phytoglycogène, un glucane soluble qui n'a pas cristallisé (**Figure 11A**). Avant la découverte des protéines ESV1 et LESV, il était admis que la cristallisation des molécules d'amylopectine se faisait spontanément par un processus physique impliquant l'autoorganisation de l'amylopectine après l'action de ISA1 et ISA2 (Feike et al., 2016), les mutants *isa* présentant une accumulation de phytoglycogène soluble et seulement quelques petites particules cristallisées (Liu et al 2023). L'implication de LESV dans ce processus a été démontrée dans l'étude des doubles mutants *isa/lesv* dans laquelle la formation des grains est altérée (Feike et al., 2016), dans des études de complémentation d'*isa* par LESV (Liu et al., 2023) et dans des études de spécificité de reconnaissance LESV/Glucanes (Osman et al., 2024).

Le phénotype de *lesv* est différent de *esv1*. Il ne présente pas d'altération de la dégradation nocturne de l'amidon mais accumule dans un nombre restreint de chloroplastes (5%) de tous petits grains insolubles et de petites particules potentiellement solubles (**Figure 11B gauche**). Le triple mutant *isa1/isa2/lesv* provoque la perte totale d'amidon insoluble au profit du phytoglycogène qui montre une action concertée entre les deux protéines dans le mécanisme de compaction des chaînes d'amylopectine dans les grains (**Figure 11B droite**). Enfin, la surexpression de LESV entraîne une polynucléation de

39

l'amylopectine qui se traduit par l'augmentation du nombre de grains d'amidon dans le chloroplaste (Figure 11C gauche). La surexpression de LESV dans le contexte de plantes mutantes *isa1/isa2* présente un phénotype analogue montrant que LESV est capable de compenser l'absence de *isa* et de cristalliser des molécules d'amylopectine polybranchées (Figure 11C droite). L'ensemble de ces données indique que LESV favorise la transition de phase de l'amylopectine vers la formation de grains solides au sein du chloroplaste. LESV et ESV1 occupent des rôles distincts mais semblent être fonctionnellement liées (Liu et al., 2023).



Figure 11. Phénotypes des grains d'amidon de la lignée sauvage et des lignées mutantes LESV et ISA. Les feuilles d'Arabidopsis thaliana ont été récoltées à la fin de la journée et les grains ont été observés par microscopie électronique à transmission. Barre à 2µm. WT : lignée sauvage. OX : lignées surexprimées. Les flèches blanches indiquent les grains insolubles tandis que les flèches rouges indiquent les grains solubles sous forme de phytoglycogène. Figure adaptée de (Liu et al., 2023).

Des résultats récents que nous avons obtenus au laboratoire montrent que si LESV et ESV1 interagissent avec l'amylopectine, seule LESV interagit avec le phytoglycogène, suggérant que LESV puisse intervenir en amont ou de manière concomitante avec ISA1/ISA2 dans l'assemblage semi-cristallin de l'amylopectine (Osman et al., 2024).

ESV1 et LESV sont des protéines non enzymatiques. L'analyse de leur structure primaire n'a pas permis d'identifier de domaine catalytique connu ni de domaine CBM (Figure 12A gauche) (Feike et al., 2016). Leurs extrémités N-terminales sont de longueurs différentes, plus courte chez ESV1 (141/426 aa) que chez LESV (309/578 aa) et ont une structure prédite désordonnée. En revanche, leurs domaines C-terminaux sont très conservés entre eux et chez leurs orthologues (Feike et al., 2016). Ce domaine contient des séquences répétées riches en résidus tryptophane et aromatiques qui alternent avec des acides aspartiques ou glutamiques. La présence de ces séquences répétées, contenant des acides aminés connus pour interagir avec des glucanes tout au long du domaine, pourrait constituer une région spécifique de la fixation de longues chaînes de glucanes (Kofler and Freund, 2006).

L'analyse du modèle AlphaFold2 (AF2)(Jumper et al., 2021) montre que le domaine C-terminal est composé d'un feuillet β de 70 Å de long, constitué de 16 brins antiparallèles (**Figure 12B**) (Liu et al., 2023). Les résidus aromatiques et acides forment quatre lignes parallèles espacées de 14 Å et alternées perpendiculaires aux brins β (deux lignes aromatiques, deux lignes acides) et dont les chaînes latérales émergent sur chaque face du feuillet (Liu et al., 2023). Cet espacement entre les alignements de résidus aromatiques correspond au diamètre d'une double hélice d'amylopectine (Buléon et al., 1998). Un modèle de complexe avec l'amylopectine, calculé à partir des modélisations AF2, montre que l'hélice s'emboîte parfaitement entre les deux lignes de résidus de tryptophane. Ce domaine riche en résidus aromatiques constitue un nouveau domaine de fixation aux glucanes, ayant une affinité spécifique pour les doubles hélices d'amylopectine (**Figure 12C**) (Liu et al., 2023).



Figure 12. Structure des protéines LESV et ESV1. (A) à gauche : Séquence des protéines ESV1 et LESV avec les domaines conservés riches en tryptophane (bleu) et riches en proline (jaune). Les protéines sont adressées au chloroplaste par un peptide de transit (cTP, vert). A droite : Modèle ab initio du domaine N-terminal de LESV calculé à partir de données SAXS à l'aide du logiciel BUNCH. Le domaine C-terminal conservé est représenté en cartoon et le modèle du domaine N-terminal est représenté sous forme de sphères (une sphère par résidu d'acide aminé) (Osman et al 2024). (B) Prédiction de structure de ESV1 et LESV par AlphaFold colorée selon l'indice de confiance de la modélisation (pLDDT). Les résidus aromatiques (Trp, Tyr et Phe en rose) forment des chaînes parallèles entre elles et perpendiculaires au feuillet β . Ces chaînes sont espacées de 14Å. Les résidus acides (Asp et Glu en jaune) forment également une chaîne de résidus acides parallèle aux chaînes aromatiques. Deux chaînes de résidus acides sont également espacées de 14Å environ. (C) Modèle proposé par bio-informatique d'une chaîne d'amylopectine alignée sur les chaînes aromatiques du feuillet β de ESV1. Figure adaptée de Osman et al., 2024 et Liu et al., 2023.

Le domaine N-terminal de ESV1 est très court. En revanche, les études SAXS réalisées sur LESV entière montrent que les régions N-terminales s'enroulent autour du domaine C-terminal, et jouent probablement un rôle dans la spécificité de reconnaissance du phytoglycogène par la protéine (**Figure 12A droite**) (Osman et al., 2024).

Nous avons pu montrer par des expériences de biochimie et de microscopie à fluorescence que la présence de ESV1 protège les grains d'amidon de la digestion par une α -amylase (Osman et al., 2024).

L'ensemble de ces résultats a permis de proposer un mécanisme d'action pour LESV et ESV1 dans la biosynthèse des grains d'amidon décrit **Figure 13** : LESV favorise la transition de phase des doubles hélices d'amylopectine de la forme soluble à la forme cristalline en interagissant avec les glucanes via son domaine riche en tryptophanes. Guidées par LESV, les doubles hélices s'arrangent parallèlement et à équidistance entre elles, entrainant la nucléation en blocs d'hélices d'amylopectine et favorisant leur cristallisation. La protéine ESV1 agirait en aval de LESV pour stabiliser la matrice des grains *de novo* afin de restreindre l'accès des activités hydrolytiques des enzymes qui dégradent les glucanes (Liu et al., 2023, Osman et al., 2024).



6. Grains d'amidon semi-cristallins

Figure 13. Proposition du mécanisme d'action de LESV et ESV1. Figure adaptée de Liu et al., 2023, Ball et al., 1996, Khalid et al., 2017.

2.7 Dégradation de l'amidon

Dans cette partie, nous décrirons les mécanismes de dégradation de l'amidon qui est illustrée dans la Figure 5D. Le taux de dégradation de l'amidon est déterminé en fonction de la durée de la nuit, optimisant ainsi l'apport en carbone pour la croissance pendant les heures d'obscurité. Par conséquent, il est lié au cycle diurne et donc à la synthèse de l'amidon. La voie de dégradation diffère selon les espèces et les organes végétaux (Zeeman et al., 2007).

Elle permet de libérer des glucanes solubles à partir de l'amidon dégradé. Ces glucanes seront convertis en une forme linéaire de MOS puis hydrolysés en maltose. Certains de ces MOS ou maltoses seront réutilisés par les enzymes de la biosynthèse de l'amidon. Certains maltoses seront exportés hors des plastes pour alimenter la plante après sa conversion en saccharose par plusieurs étapes. La dégradation implique un large éventail d'enzymes : des β et α -amylases (BAM et AMY), des DBE (principalement les isoamylases ISA3) et des enzymes de phosphorylation et déphosphorylation (a-Glucan water dikinase (GWD), phosphoglucan water dikinase (PWD), Starch Excess 4 (SEX4) et Like Starch Excess Four 2 (LSF2).

La dégradation des grains d'amidon débute par l'action des dikinases GWD et PWD. Elles transfèrent le β-phosphate de l'ATP sur des résidus de glucoses de l'amylopectine aux positions C6 et C3 pour permettre aux protéines AMY, BAM et ISA d'accéder aux doubles hélices d'amylopectine déstabilisées par l'ajout des groupements phosphate (Monroe and Storm, 2018; Ritte et al., 2006; Streb et al., 2009; Zeeman et al., 2004). Ensuite, les BAM, des enzymes exo-hydrolytiques, coupent une liaison α-1,4 sur deux à partir de l'extrémité non réductrice pour libérer du β-maltose (Kaplan et al., 2006). Arabidopsis possède neuf gènes de β-amylases (BAM1 à 9) (Fulton et al., 2008). L'enzyme de déramification, l'isoamylase, intervient quant à elle pour hydrolyser les points de ramification, α -1,6 et génère de longues chaînes linéaires d'amylopectine, qui seront converties en maltose et maltotriose (Pfister and Zeeman, 2016; Streb et al., 2009). En parallèle, les chaînes d'amylopectine déstabilisées par phosphorylation sont également déphosphorylées par d'autres enzymes afin de les rendre accessible aux enzymes de dégradation. Les glucanes sont ainsi déphosphorylés en positions C6 et C3 par SEX4 et en position C3 par LSF2 (Kötting et al., 2009; Santelia and Zeeman, 2011). Cette alternance de phosphorylation/déphosphorylation permet de déstabiliser la surface du grain et le rendre accessible aux hydrolases.

Enfin, Les MOS, maltose et maltotriose seront soit repris pour la synthèse d'un nouveau grain d'amidon soit pourront être exportés par des transporteurs spécifiques tels que le transporteur de maltose 1 (MEX1) hors du chloroplaste dans le cytosol pour la conversion en saccharose.

Chapitre 3. Initiation de la biosynthèse de l'amidon

3.1 SS4 est une enzyme cruciale pour l'initiation des grains d'amidon

*At*SS4 a été caractérisée pour la première fois par (Roldán et al., 2007) comme étant l'enzyme majeure de l'initiation, permettant de créer des précurseurs glycaniques qui pourront être utilisés par les autres SSs (**Figure 5B**). Des études de génétique inverse sur les mutants *ss4* (Columbia-0 et Wassilewskija) ont permis d'avancer dans la compréhension de sa fonction. Ces deux mutants présentent des phénotypes altérés de la plante sauvage (WT ; type sauvage) au niveau du nombre, de la morphologie des grains d'amidon et de la croissance de la plante.

3.1.1 SS4 est impliquée dans la taille, la forme et l'initiation de la biosynthèse des grains d'amidon

La première différence phénotypique de *ss4* par rapport au WT s'observe sur la croissance de la plante. *ss4* présente un retard de croissance qui se traduit par des feuilles, appelées rosettes, plus petites et plus pâles (Roldán et al., 2007) (**Figure 14B**). Ce retard de croissance s'explique par une accumulation d'ADPG qui n'a pas été utilisé pour la synthèse amidon. L'accumulation d'ADPG dans le mutant *ss4* suggère que les autres SSs ne peuvent pas l'utiliser pour la synthèse du grain. Ce phénotype montre l'importance de la SS4 pour la croissance normale de la plante. Cela met en évidence que cette isoforme est impliquée dans l'étape d'initiation en amont de l'action des autres SSs dans la synthèse du grain.

La deuxième différence montre l'importance de la SS4 dans le contrôle du nombre du grain d'amidon dans le plaste. Les chloroplastes des rosettes matures de *ss4* contiennent 0 seul voire 1 grain par plaste contre 4 à 7 grains dans le WT (Roldán et al., 2007) (**Figure 14A**) et les jeunes rosettes ne contiennent quant à elles pas d'amidon (Crumpton-Taylor et al., 2013). De plus les rares grains observés au microscope électronique à transmission présentent une cavité en zone centrale, évoquant une défaillance dans l'initiation de la synthèse du grain. Cette baisse du nombre de grains n'est *a priori* pas due à l'accumulation d'ADPG. Premièrement, l'exposition à la lumière continue des plantes mutantes *ss4* a permis de restaurer le nombre de grains presque à la normale mais aucune diminution du taux d'ADPG n'a été observée (Malinova et al., 2017). Deuxièmement, dans les doubles mutants *ss4/dpe2* (Disproportionating enzyme) ou *ss4/phs1* (α-glucan phosphorylase), protéines impliquées dans la conversion du maltose et dans la dégradation phosphorolytique des chaînes de glucose respectivement, on observe une diminution du nombre de grains par rapport au WT sans qu'il y ait d'accumulation d'ADPG (Malinova et al., 2018, 2017). PHS1 a été décrite avec un rôle possible dans le métabolisme des MOS au sein des plastes et contribue à la fois dans la synthèse et la dégradation de l'amidon. SS4 interagirait par ailleurs avec PHS1 mais la fonction de cette interaction pour la synthèse des grains n'est pas bien comprise (Malinova et al., 2018).

La taille et la forme des grains d'amidon de ss4 sont aussi modifiées par rapport au WT. Les grains ss4 sont plus gros, passant de 1-2 µm pour le WT à 3-5 µm et sont gonflés alors qu'ils sont de forme lenticulaire dans le WT (Roldán et al., 2007; Vandromme et al., 2019)(**Figure 14A et Figure 16A**). Bien que, ces changements de taille puissent être liés à l'espace disponible dans le chloroplaste, la morphologie est néanmoins bien définie par SS4 (Crumpton-Taylor et al., 2012; Crumpton-Taylor et al., 2013). En effet, cet aspect gonflé du grain n'est pas retrouvé dans d'autres mutants impliqués dans l'initiation (*pii1,mfp1,ptst2 et ptst3*). Ces mutants seront décrits dans la partie 3.2. SS4 influence donc également la morphologie du grain.

Enfin, la comparaison de *ss4* avec les mutants des autres isoformes de SSs renforce l'idée que SS4 agit pendant l'initiation. Contrairement à *ss4*, les mutants *ss1*, *ss2 et ss3* ne présentent pas de changement de forme et du nombre de grains mais présentent, en revanche, une altération de la structure de l'amylopectine qui n'est pas observée dans le mutant *ss4*.



Figure 14. Phénotypes des grains et rosettes de la plante sauvage et du mutant ss4. (A) Phénotypes des grains d'amidon de chloroplastes de feuilles sauvages (Columbia-0) et mutants ss4 par microscopie électronique à balayage. Les flèches jaunes indiquent la position des grains d'amidon. Figure adaptée de Seung et al 2018. *(B)* Rosettes colorées à l'iode des feuilles (Columbia-0) et Atss4. Atss4, accumule moins de grains d'amidon en comparaison à la lignée sauvage. Figure adaptée de Seung et al., 2018.

3.1.1 SS4 n'est pas la seule enzyme capable d'initier

En l'absence de SS4, on observe encore dans certains chloroplastes la présence d'un grain montrant que l'initiation n'est pas conditionnée uniquement par la présence de SS4. Le double mutant *ss3ss4* montre un phénotype encore plus aberrant que le simple mutant *ss4* avec une absence presque totale de grains par chloroplaste et la croissance de la plante encore plus affectée que dans le mutant *ss4* (Szydlowski et al., 2009). SS3 serait donc capable d'initier la synthèse des grains de façon partielle en utilisant l'ADPG mais ne permet pas autant d'évènements d'initiation que SS4. Cette enzyme pourrait être responsable de l'initiation des quelques grains encore observés dans le mutant *ss4*.

3.1.2 SS4 génère des amorces pour les autres amidon synthases

SS4 est la première enzyme à intervenir dans la biosynthèse du grain parmi les SSs, ce qui suggère que celles-ci ne sont pas capables d'utiliser les amorces de court DP générées *de novo* dans le chloroplaste. A partir de ces amorces, SS4 va générer une préamylopectine reconnaissable par les autres SSs. Cependant, le substrat exact de SS4 fait encore débat.

Il existe des points d'incompréhension concernant l'origine des MOS dans le chloroplaste. La formation des MOS peut prendre son origine de deux façons selon que les chloroplastes possèdent ou non des grains déjà synthétisés. Lorsque la plante possède des grains déjà présents, les MOS proviennent pour la plupart de la dégradation des grains, par l'action des DBE et des amylases. La dégradation des grains crée un pool important de MOS réutilisables par les SSs. Cependant, lorsqu'il n'y a pas encore eu d'initiation dans un chloroplaste exempt de grain, la synthèse *de novo* des MOS n'est pas connue.

Chez les mammifères, le glycogène nécessite la présence de la glycogénine, pour être synthétisé. Cette glycosyltransférase a la particularité de s'auto-glucosyler et de se lier de manière covalente avec un court oligosaccharide ((Barengo et al., 1975; Lomako et al., 1988). Chez les plantes, il n'existe pas de preuve de la présence de fonction de type glycogénine qui pourrait initier la synthèse de l'amidon.

Le maltose a été décrit comme étant le plus petit accepteur possible utilisé par les SSs, là où le glucose et l'ADPG ne peuvent pas être utilisés comme tels (Brust et al., 2014). Il pourrait alors servir d'amorce pour l'initiation.

3.1.4 Rôles distincts des domaines de SS4 dans l'initiation

SS4 est une protéine de 117 kDa composée de 1040 acides aminés, avec un domaine N-terminal (acide aminé 1 à 466) et un domaine C-terminal (467 à 1040) (**Figure 15**).



Figure 15. Structure prédite de SS4. SS4 comporte un domaine N-terminal composé d'un peptide de transit (*TP*, vert), de quatre régions prédites structurées en coiled-coils (orange) et d'un domaine de dimérisation (violet, *RD*). Le domaine C-terminal contient le domaine catalytique composé de deux domaines glycosyl transferases 1 et 5 (*GT*1 et *GT*5). Selon Raynaud et al., 2016.

Le domaine catalytique de SS4, très conservé parmi les SSs, contient un site actif situé entre les domaine GT1 et GT5 (Cf chapitre 2.3) (Lombard et al., 2014; Nielsen et al., 2018). Il permet le transfert d'une molécule de glucose issue de l'ADPG vers l'extrémité non réductrice d'une chaîne linéaire par un mécanisme de rétention de configuration.

Le domaine N-terminal de *At*SS4 n'est pas conservé parmi les SSs mais conservé parmi les orthologues de SS4 (Lu et al., 2018). Il contient un peptide de transit de 42 résidus chez *Arabidopsis*. De plus, Raynaud et al., 2016 ont décrit quatre régions CC prédites et une région qui servirait à la dimérisation de SS4 (RD) d'environ 50 acides aminés, conservée parmi les orthologues et requise pour la formation des homodimères (**Figure15**) (Raynaud et al., 2016).

Le domaine N-terminal est impliqué dans la détermination de la morphologie des grains d'amidon dans le chloroplaste. Les grains d'amidon ont été observés pour deux lignées transgéniques exprimant soit uniquement le domaine N-terminal de SS4 (SS4-Nter) soit uniquement le domaine C-terminal (SS4-Cter) fusionnés avec une protéine fluorescente (YFP) (**Figure 16**) (Lu et al., 2018). Les grains de la lignée SS4-Cter sont plus gros mais tout aussi nombreux que la lignée sauvage. Cependant, les grains présentaient une forme arrondie, caractéristique du mutant *ss4*. Les grains des lignées SS4-Nter présentent un phénotype (1 seul gros grain arrondi) similaire à la lignée *ss4*, montrant que d'une part la région N-terminale n'a pas d'effet sur la biosynthèse et d'autre part qu'elle influence la forme du grain (Lu et al., 2018). Ces travaux montrent également le rôle du domaine C-terminal catalytique dans le contrôle du nombre de grains par chloroplaste. Plus important encore, ces travaux indiquent que la morphologie du grain d'amidon est définie pendant l'étape d'initiation.



Figure 16. Phénotypes des grains d'amidon des lignées sauvage SS4 et ss4. Adaptée des résultats de Lu et al., 2018.

La région N-terminale est nécessaire pour la localisation correcte de SS4 dans le chloroplaste. SS4 est localisée sous forme de spots au bord des grains d'amidon et non pas dans le stroma (Gámez-Arjona et al., 2014; Szydlowski et al., 2009). Lorsque la partie N-terminale seule de SS4 est exprimée, la localisation n'est pas affectée tandis que dans SS4-Cter la protéine est présente de façon diffuse dans le stroma du chloroplaste (Lu et al., 2018). Il a été montré récemment que la localisation de SS4 est due à son interaction avec une autre protéine non enzymatique (MFP1) localisée dans la membrane des thylakoïdes. Le rôle de MFP1 sera décrit dans la partie suivante.

3.2 Un complexe d'initiation impliquant des interactions protéineprotéine

3.2.1 SS4 interagit avec plusieurs protéines dans le chloroplaste

Le rôle de SS4 à lui seul ne suffit pas pour expliquer la fréquence et la localisation des mécanismes d'initiation. La présence de CC a conduit à l'hypothèse selon laquelle la protéine SS4 aurait des partenaires d'interaction protéique. Pour tenter de mieux comprendre ces mécanismes, des recherches de partenaires ont été entreprises.

Dans un premier temps, deux homologues de PTST1 dans le génome d'Arabidopsis, PTST2 et PTST3 ont été identifiés par immunoprécipitation en utilisant SS4 comme protéine marquée (Seung et al., 2017). Ces deux protéines servent à l'initiation correcte des grains dans les chloroplastes et sont prédites avec un domaine CC accolés à un CBM48 permettant de se lier aux MOS. Dans un deuxième temps, deux autres protéines non enzymatiques ont été identifiées par immunoprécipitation avec PTST2 comme protéine marquée ou par criblage en double hybride avec SS4 comme protéine appât (Seung et al., 2018; Vandromme et al., 2019). Il s'agit des protéines PII1 et MFP1. Ces deux protéines ont été désignées respectivement comme permettant la régulation des évènements d'initiation et la localisation aux membranes de thylakoïdes. Ces protéines ne contiennent ni domaine catalytique, ni domaine de fonction connue mais présentent des régions CC prédites.

En l'absence de PTST2, PTST3, MFP1 ou PII1, les chloroplastes n'accumulent presque aucun grain d'amidon (Seung et al., 2018; Vandromme et al., 2019). Ces recherches révèlent que la formation d'un complexe multiprotéique est essentielle à l'initiation correcte des grains d'amidon.

La **Figure 17** résume les caractéristiques physiques des grains issues des mutants d'Arabidopsis thaliana en comparaison avec la lignée sauvage. Chaque mutant sera décrit dans le chapitre correspondant.

















Figure 17. Phénotypes des grains d'amidon des lignées mutantes et doubles mutantes des protéines intervenant dans l'initiation. Les chloroplastes sont observés sur des sections de feuilles colorées au bleu de toluidine. Barre à 10µm. Les grains sont observés au microscope optique à l'exception des mutants ss3, pii1ss3 et pii1ss4 qui sont observés en microscopie confocale. Figure adaptée de Vandromme et al., 2024, Abt et al., 2020, Seung et al., 2017, 2018 et Crumpton-Taylor et al., 2016

3.2.2 PTST2 apporte les MOS à SS4 et interagit avec PII1

PTST2 et PTST3 ont été identifiées par des analyses phylogénétiques à partir de la séquence de PTST1 (Seung et al., 2017). PTST2 est conservée parmi les autres espèces. Il existe un orthologue de PTST2, FLO6, chez le riz et l'orge connu pour être impliqué dans l'initiation. La mutation *flo6* provoque une diminution du nombre de grains d'amidon. Dans l'albumen des graines, FLO6 interagit également avec ISA1 mais cette interaction n'est pas retrouvée chez *Arabidopsis* (Peng et al., 2014; Saito et al., 2018; Seung et al., 2017).

PTST2 et PTST3 possèdent un CBM48 et une région CC (**Figure 18**). Il a été montré qu'avec son CBM, PTST2 interagit préférentiellement avec des MOS qui ont tendance à adopter une structure secondaire hélicoïdale (au moins 10 unités de glucose) (Seung et al., 2017). PTST2 pourrait se lier principalement aux doubles hélices d'amylopectine. L'interaction entre PTST2 et SS4 a été vérifiée par immunoprécipitation sur des extraits de feuilles (Seung et al., 2017). Cependant, un nombre très faible de peptides de SS4 a été reconnu dans l'interaction ce qui pourrait signifier que SS4 interagirait de façon transitoire avec PTST2 au cours de l'initiation (Seung et al., 2017). PTST2 semble indispensable à l'initiation du nombre de grains car le mutant *ptst2* ne contient qu'un seul grain par chloroplaste comme le mutant *ss4* (**Figure 17**). En revanche, la forme lenticulaire de ces grains et la croissance de la plante restent inchangées par rapport à la plante sauvage.

Ainsi, on pense que PTST2 amène le substrat à la SS4 (qui ne possède pas de CBM). Cependant, si PTST2 se lie aux doubles hélices d'amylopectine, même courtes, des questions subsistent quant à la façon dont SS4 initie la synthèse du grain lorsqu'aucune double hélice n'est présente (Seung et al., 2018). Ces doubles hélices associées à SS4 par l'intermédiaire de PTST2 pourraient faciliter la cristallisation des chaînes de glucanes au niveau du hile du grain (Seung and M. Smith, 2018).



Figure 18. Structure prédite de PTST2 et PTST3. En vert : peptide de transit (TP). En bleu : domaines coiled-coils prédits (CC). En violet : module de liaison aux sucres (CBM48). Figure adaptée de Seung et al., 2018.

PTST2 et PTST3 interagissent également ensemble mais contrairement à PTST2, PTST3 n'est pas conservée chez les graminées et son rôle dans la plante serait partiellement redondant à celui de PTST2 (Seung et al., 2018). Le mutant *ptst3* présente un phénotype moins marqué que celui du mutant *ptst2*, avec une réduction plus faible du nombre de grains d'amidon, généralement deux grains par chloroplaste (Seung et al., 2018). Par contre, le double mutant *ptst2/ptst3* contient moins de grains que dans les mutants simples ce qui semble montrer que les deux protéines sont importantes pour l'initiation et que PTST3 peut compenser partiellement la perte de PTST2 (**Figure 17**).

Enfin, PTST2 a été identifiée en interaction avec deux autres partenaires : PII1 et MFP1. L'interaction avec PII1 a été révélée par immunoprécipitation avec un plus grand nombre de peptides identifiés par rapport à l'interaction entre PTST2 et SS4, ce qui pourrait signifier une interaction plus stable (Seung et al., 2018). Cependant, la fonction de cette interaction au sein du complexe d'initiation n'est pas connue. Enfin, PTST2 nécessite la présence de MFP1 qui établit l'emplacement correct de PTST2 dans le chloroplaste.

3.2.3 MFP1 localise le complexe aux thylakoïdes

La MAR-binding Filament like Protein 1 (MFP1) est décrite comme étant une protéine non enzymatique qui facilite la localisation des protéines d'initiation au niveau des membranes des thylakoïdes (Sharma et al., 2024). MFP1 est exclusivement localisée du côté stromal des membranes des thylakoïdes (Jeong, 2003). PTST2 et MFP1 co-localisent dans des localisations spécifiques au niveau des thylakoïdes (Seung et al., 2018). L'extrémité N-terminale de MFP1 contient un peptide de transit prédit pour sa localisation dans le chloroplaste. De plus elle contient une séquence signal de translocation aux thylakoïdes twin arginine (TAT) suivie d'un court domaine transmembranaire (DT) (**Figure 19**). L'extrémité C-terminale contient une longue région CC prédite. La séquence TAT permet d'importer la protéine aux thylakoïdes et sa région transmembranaire permettrait de s'ancrer dans les membranes des thylakoïdes avec la région CC positionnée face au stroma, facilitant les interactions avec les partenaires (Sharma et al., 2024; Xu et al., 2017).



Figure 19. Structure prédite de MFP1 impliquée dans l'initiation. Figure adaptée de Sharma et al., 2024.

Dans le mutant *mfp1*, le nombre de grains d'amidon par chloroplaste est réduit à 1 ou 2 grains par chloroplaste comparé à 4 à 7 dans la lignée sauvage, et ces grains sont plus gros que ceux observés dans la plante sauvage (**Figure 17**). Cependant, la teneur en amidon reste comparable à celle de la plante sauvage (Seung et al., 2018). De plus PTST2 est associée aux thylakoïdes et dans le stroma sous forme de spots. La mutation *mfp1* provoque la délocalisation de PTST2 exclusivement dans le stroma (Seung et al., 2018). MFP1 servirait à guider PTST2 à des endroits spécifiques et par conséquent le complexe d'initiation (Sharma et al., 2024).

Pour confirmer le rôle de MFP1, un variant dépourvu du peptide signal au thylakoïde et du domaine transmembranaire (MFP1ΔTAT-DT), a été exprimé (Sharma et al., 2024). L'objectif était de déterminer si le DT est nécessaire pour la localisation de MFP1 aux membranes des thylakoïdes. Ce variant a été détecté grâce à une fusion avec l'étiquette YFP (protéine fluorescente jaune). Comme prévu, MFP1ΔTAT-DT a été localisé dans la fraction soluble plutôt que dans la fraction membranaire. MFP1ΔTAT-DT a ensuite été exprimée en présence de la protéine Tic40, qui se localise à la membrane interne de l'enveloppe des chloroplastes. L'objectif était de vérifier si Tic40 modifie la localisation de la synthèse d'amidon dans le chloroplaste. De manière surprenante, la microscopie électronique à balayage a montré des grains d'amidon localisés au niveau de la membrane chloroplastique plutôt qu'accolés aux thylakoïdes (Sharma et al., 2024). Cela indique que la localisation de MFP1 est suffisante pour rediriger la majorité (voire l'ensemble) du complexe d'initiation vers un compartiment sous-chloroplastique (Sharma et al., 2024).

Ensuite, pour vérifier si TIC40-MFP1- Δ TAT-DT est suffisant pour relocaliser PTST2 à l'enveloppe chloroplastique, PTST2-mcitrine a été intégré au double mutant *mfp1/ptst2* en présence de MFP1 ou MFP1- Δ TAT-DT-TIC40 (Sharma et al., 2024). Dans le premier cas PTST2 est localisé sous forme de spot au sein du chloroplaste tandis que lorsque PTST2 est exprimée avec TIC40-MFP1- Δ TAT-DT, les spots sont localisés à la périphérie du chloroplaste. La même expérience a été réalisée pour localiser SS4 en utilisant les mutants *ss4* ou *mfp1/ss4*. Dans le premier cas SS4-mcitrine est localisé dans le stroma sous forme de spots discrets tandis que lorsque SS4 est relocalisé à la périphérie des chloroplastes exprimée dans le fond mutant *mfp1/ss4* en présence de TIC40-MFP1- Δ TAT-DT. Cela démontre que MFP1 contrôle la localisation de PTST2 directement et indirectement de SS4 via PTST2 (Sharma et al., 2024).

Enfin, l'expression de MFP1ΔTAT-DT associée à la protéine FTSH5 Interacting Protein 1 (FIP1), qui se localise aux thylakoïdes, a permis de restaurer la localisation de MFP1 aux membranes des thylakoïdes et de rétablir le nombre de grains d'amidon par rapport à la lignée *Atmfp1*. Ces résultats suggèrent que l'initiation des grains se déroule au niveau de la membrane des thylakoïdes, et que MFP1 est impliquée dans la localisation du complexe d'initiation en interagissant avec PTST2.

En résumé, MFP1 est impliquée dans l'initiation des grains d'amidon en facilitant la localisation des protéines d'initiation aux membranes des thylakoïdes. En tant que protéine non enzymatique, MFP1 interagit principalement avec PTST2 plutôt qu'avec SS4, et cette interaction est essentielle pour la formation normale des grains.

57

3.2.4 PII1 contrôle le nombre d'évènement d'initiation

La Protein Involved in Initiation 1 (PII1) joue un rôle central dans la régulation de l'initiation des grains d'amidon, en interagissant avec SS4, PTST2 et SS5. Identifiée comme un interactant direct de SS4 par criblage double hybride (Vandromme et al., 2019) et par immunoprécipitation avec PTST2 (Seung et al., 2018), PII1, également connu sous le nom de MRC (Myosin Ressembling Chloroplast Protein), est un régulateur clé dans la formation des grains d'amidon. PII1 est hautement conservée parmi les génomes végétaux supérieurs et dans tous les tissus photosynthétiques (Schmid et al., 2005).

PII1 se distingue de PTST2 et SS4 par la présence de CC prédits sur la quasi-totalité de sa structure, ce qui suggère qu'elle pourrait jouer un rôle dans la formation de complexes protéiques. En revanche, aucun domaine fonctionnel n'a été identifié, lui préférant plutôt un rôle de régulation ou mécanique (**Figure 20**).



100 a.a.

Figure 20. Structure prédite de PII1 impliquée dans l'initiation. En vert : peptide de transit (TP). En bleu : domaines coiled-coils prédits (CC). Figure adaptée de Seung et al., 2018.

Le mutant *pii1* présente un phénotype similaire à celui du mutant *ss4*, avec un seul grain d'amidon par chloroplaste, et de taille augmentée (2,5-3,3 μ m) mais de forme lenticulaire (**Figure 17**). Cependant, contrairement à *ss4* ou *ptst2*, chaque chloroplaste contient au moins un grain ce qui indique que la synthèse d'amidon est moins gravement affectée. Le mutant *pii1* accumule peu d'ADPG dans les feuilles ce qui montre qu'en l'absence de PII1 il peut être utilisé pour la synthèse d'amidon. Comme dans les mutants *ptst2* ou *mfp1*, la croissance des plantes et la forme des grains d'amidon restent comparables à celles du type sauvage et le ratio amylose/amylopectine et la teneur en amidon restent inchangés (Vandromme et al., 2019).

Récemment SS5 a été décrite comme étant impliquée dans l'initiation en interaction avec PII1 (Abt et al., 2020). Les mécanismes biochimiques par lesquels, SS4 ou SS5 interagissent

avec PII1 ne sont pas compris. Le double mutant *pii1/ss4* présente un phénotype plus sévère que le simple mutant *ss4*, certains chloroplastes ne contenant plus aucun grain d'amidon, tandis que *pii1/ss5* montre un phénotype moins sévère que le mutant *ss4* (**Figure 17**). Cela indique qu'en l'absence de PII1, SS4 peut initier un grain d'amidon par chloroplaste, mais pas plusieurs. Cette observation s'appuie sur l'hypothèse selon laquelle PII1 contrôlerait le nombre d'événements d'initiation des grains via SS4 (Vandromme et al., 2019). L'interaction entre MFP1 et PII1 n'a pas été montrée mais le mutant *pii1/mfp1* présentait également un nombre plus élevé de chloroplastes sans grain en comparaison de la lignée *pii1* (Seung et al., 2018). PII1 est localisée sous forme de spots dans la fraction stromale (Vandromme et al., 2019, Seung et al., 2018) ce qui est probablement dû à son interaction avec PTST2.

La protéine SS5 est localisée, comme PII1 et SS4, sous forme de spots dans un espace sous-chloroplastique (Seung et al., 2018 ; Abt et al., 2020). L'expression d'*At*SS5 dans un système de biosynthèse transposé chez la levure a montré que SS5 n'allonge pas les glucanes, contrairement à ses homologues SS1, SS2, SS3, et SS4 (Abt et al., 2020). La structure de SS5 montre son rôle probable dans la régulation du nombre de grains d'amidon. SS5 possède, comme les autres SSs, un domaine N-terminal constitué d'hélices et un domaine C-terminal globulaire ne contenant que le domaine GT5 donc dépourvu du domaine de fixation à l'ADPG et par conséquent inactif. SS5 est décrite comme une protéine dimérique (Abt et al., 2020). La suppression de son CC en N-terminal empêche la dimérisation de SS5 et son interaction avec PII1, soulignant l'importance du domaine N-terminal dans l'initiation. Helle et al., 2018 ont suggéré que SS5 chez *Solanum tuberosum (St)* pourrait interagir avec les glucanes via ses régions CC. Bien que SS5 soit une protéine non canonique, car elle est dépourvue d'activité glycosyltransférase, elle conserve des résidus aromatiques impliqués dans la liaison au glucane, contrairement à l'acide aspartique essentiel à cette interaction chez SS4.

En résumé, le mécanisme par lequel PII1 interagit avec SS4 et SS5 et d'autres protéines comme PTST2 reste à clarifier, mais il est clair que ces interactions sont essentielles pour la régulation efficace au démarrage de la formation des grains d'amidon.

59

3.2.5 Résumé et connaissances actuelles sur l'initiation

Au cours du mécanisme d'initiation, SS4 est l'acteur principal qui permet la formation d'une amorce pré-amylopectine. SS3 peut également permettre l'initiation, mais son rôle est principalement associé à l'allongement des chaînes de glucose. La protéine SS5 est également impliquée dans l'initiation mais ne possède pas d'activité propre d'amidon synthase et son rôle est discuté. De plus, quatre protéines non enzymatiques jouent chacune un rôle précis dans ce processus. MFP1 permet de localiser le complexe multiprotéique aux membranes des thylakoïdes par l'intermédiaire de PTST2 pour concentrer spécifiquement l'initiation des grains entre les membranes des thylakoïdes. PTST2 quant à elle servirait à faciliter l'initiation en fournissant les MOS longs à SS4 et enfin PII1 interviendrait pour réguler les évènements d'initiation, en interaction avec PTST2, SS4 et SS5. Toutes ces protéines sont impliquées dans un complexe d'initiation multiprotéiques (Figure 21).



Figure 21. Interactions des protéines intervenant dans l'initiation.

L'apparition de nouveaux acteurs impliqués dans un complexe d'initiation soulève des questions concernant la dynamique spatio-temporelle des interactions protéineprotéine. À ce stade, nous ne savons pas si les interactions entre MFP1/PTST2, PTST2/SS4 ou bien SS4/PII1 se font de façon transitoire ou s'il s'agit d'un complexe multiprotéique stable. Le contrôle du nombre d'événements d'initiation au sein du chloroplaste est également mal connu. Une autre ambiguïté réside aussi dans le fait que l'initiation diffère également selon l'organe et les espèces végétales. Par exemple, alors que les feuilles d'*Arabidopsis* contiennent une population uniforme de grains lenticulaires, l'albumen de blé contient deux populations (type A et B) de grains ronds qui résulte de plusieurs évènements d'initiation (Seung and M. Smith, 2018). Dans les feuilles d'*Arabidopsis*, PII1 favorise l'initiation des grains tandis que dans l'albumen de blé, PII1 jouerait plutôt un rôle antagoniste en réprimant l'initiation des grains de type B en début de développement (Thèse de Jiawen Chen 2022). Cela montre le rôle complexe de PII1 en tant que régulateur du mécanisme d'initiation.

Enfin, il est probable que le métabolisme de l'amidon soit régi par de multiples contrôles post-traductionnels (phosphorylation) ou transcriptionnels. La synthèse et la dégradation sont coordonnées selon le cycle jour/nuit pour s'adapter aux changements environnementaux. En cas de changement de luminosité, température ou de stress oxydatif, la synthèse des grains est affectée. Enfin, l'initiation est également liée à l'étape de dégradation qui fournit le pool de MOS nécessaire à la synthèse de nouveaux grains. Cependant lorsque la plante n'a pas encore produit les premiers grains par chloroplaste, Le mécanisme par lequel le complexe d'initiation utilise les MOS disponibles reste encore mal compris.



Les protéines SS4 et PII1 sont des protéines clés pour l'initiation mais le mécanisme par lequel elles interagissent ensemble n'est pas clair. SS4 et PII1 possèdent toutes deux des régions CC prédites qui pourraient servir à cette interaction. Le rôle plutôt mécanique ou de régulation que PII1 exerce sur SS4 non plus n'est pas bien compris.

Le but de ce travail de thèse est de caractériser la fonction de PII1 et, au-delà, de progresser dans la connaissance du mécanisme d'initiation des grains d'amidon par l'étude du rôle de l'interaction PII1/SS4. Pour cela J'ai pour objectif d'exprimer les deux protéines et d'étudier leur structure ainsi que celle du complexe PII1/SS4 par des approches de biophysique telles que le dichroïsme circulaire utilisé en routine au laboratoire et associé à la modélisation moléculaire et l'étude enzymatique.

Le projet se divise en trois tâches principales dont les résultats sont présentés en quatre chapitres :

Chapitre 1,2 et 3 : Production et purification de PII1, SS4 et PII1/SS4

Chapitre 4 : Étude de la structure de PII1 et SS4 par des approches de biophysique et de bioinformatique ;

Chapitre 5 : Étude de la fonction enzymatique de SS4 seule et en complexe avec PII1.

MATERIEL ET METHODES

Chapitre 1. Matériels et solutions utilisés

1.1 Souches bactériennes

Plusieurs souches bactériennes ont été utilisées dans ce travail et sont énumérées ciaprès :

Top10 (Invitrogen[™]) : souche *d'Escherichia coli* (*E.coli*) conçue pour une grande efficacité de transformation. Nous avons utilisé cette souche lors des clonages pour la préparation des plasmides. Genotype : F- mcrA D(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara- leu)7697 galU galK rpsL (StrR) endA1 nupG

BL21(DE3) (Invitrogen[™])): souche *d' E.coli* optimisée pour la production de protéines recombinantes. Genotype : F – ompT hsdSB (rB- mB-) gal dcm (DE3)

BL21-AI (Invitrogen™) : souche d'E.coli permettant l'expression des protéines toxiques. Ces souches sont inductibles par addition de L-arabinose dans le milieu de culture. Génotype :
F- ompT hsdSB(rB- mB-) gal dcm araB ::T7RNAP-tetA

Rosetta-Gami2 (DE3) (Invitrogen[™])) : souche dérivée de *E.coli* BL21(DE3) et optimisée pour la reconnaissance des codons rares chez Escherichia coli et l'expression cytoplasmique de protéines contenant des ponts disulfures. Genotype : Δ(ara-leu)7697 ΔlacX74 ΔphoA Pvull phoR araD139 ahpC galE galK rpsL (DE3) F'[lac+lacIq pro] gor522::Tn10 trxB pRARE2 (CamR, StrR, TetR)

C41(DE3) (Over Express[™]) : souche *d' E.coli* contenant au moins une mutation qui permet l'expression de protéines toxiques. Genotype : F – ompT hsdSB (rB- mB-) gal dcm (DE3)

C43(DE3) (Over Express[™]) : souche *d' E.coli* dérivée de C41(DE3) possédant plusieurs mutations permettant la production de protéines toxiques différentes de celles exprimées avec C41(DE3). Genotype : F – ompT hsdSB (rB- mB-) gal dcm (DE3

1.2 Vecteurs

Plusieurs vecteurs d'expression ont été utilisés pour le clonage et l'expression des protéines étudiées dans ce travail.

pENTR/D-TOPO : vecteur d'entrée Gateway[™] permettant le clonage directionnel des gènes d'intérêt dans les vecteurs d'expression.

pET300Nt : vecteur de destination Gateway[™] facilitant la purification grâce à l'ajout d'une étiquette 6-Histidine (6-His) à l'extrémité N-terminale des protéines exprimées.

pML310 : vecteur de destination Gateway[™] facilitant la purification grâce à l'ajout d'une étiquette 6-His et la thioredoxine (TRX) à l'extrémité N-terminale des protéines exprimées.

pML333 : vecteur de destination Gateway[™] facilitant la solubilisation et la purification grâce à l'ajout d'une étiquette 6-His et une protéine de liaison au maltose (MBP) à l'extrémité Nterminale des protéines exprimées.

1.3 Solutions d'usage courant

La composition des solutions utilisées dans ce travail est décrite ci-dessous :

Tampon de migration pour électrophorèse en gel de polyacrylamide dénaturant (SDS-PAGE) : Tris-HCl 50 mM, pH 8.5 ; glycine 200 mM ; SDS 0,1 % (v/v).

Tampon de migration pour électrophorèse en gel de polyacrylamide non dénaturant (PAGE) : Tris-HCl 50 mM, pH 8.5 ; glycine 200 mM.

Tampon de migration pour électrophorèse en gel d'agarose : Tris 40 mM ; acétate de sodium 5 mM, pH 8.0 ; EDTA 1 mM.

Mélange d'incubation des zymogrammes permettant d'étudier l'activité enzymatique d'élongation de chaînes de glucanes: 67 mM Glycyl-glycine/NaOH pH 9 ; 133 mM sulfate d'ammonium ; 80 mM MgCl2 ; 0,6 mg/ml sérum-albumine bovine ; 25 mM βmercaptoéthanol ; 2,4 mM ADPG. Tampon de transfert pour Western-Blot : Tris-glycine 1X, méthanol 20%.

Tampon de lavage pour Western-Blot (TBST) : 0,05 mol/L de Tris-HCl, 0,3 mol/L de NaCl, 0.1% Tween® 20.

Solution de coloration des gels de polyacrylamide : Bleu de Coomassie (InstantBlue®, Sigma-Aldrich).

Bleu de charge des gels dénaturants (10X) : Tris-Glycine 1X, SDS 0,1X, Glycérol 25%, EDTA 10 mM, β-mercaptoéthanol 0,5%.

Bleu de charge des gels non dénaturants (10X) : Tris-Glycine 1X, Glycérol 25%, EDTA 10 mM.

Bleu de charge des gels d'agarose (6X) : 100 mM tris-HCL, pH 8,0, Xylène Cyanol 0,03%, Bleu de Bromophenol 0,03%, Glycérol, 30 %, 60 mM EDTA, pH 8,0

Marqueur de taille des gels d'agarose : SmartLadder™ de 200 à 10000 bp (EurogenTec).

Marqueur de taille des gels de polyacrylamide : PageRuler™ de 10 à 250 kDa (Thermo Fisher)

Tampon TFBI: CH₃CO₂K 30 mM, CaCl₂ 10 mM, MnCl₂ 50 mM, RbCl 100 mM, glycérol 15 % (v/v), pH 5,8.

Tampon TFBII: MOPS 10 mM, CaCl. 75 mM, RbCl 100 mM, glycérol 15% (v/v), pH 6,5.

20X NPS (1L): 66g (NH4)2SO4, 136 g KH2PO4, 142 g NA2HPO4.

50x5052 (1L): 250 g glycérol, 25 g glucose, 100 g $\,\alpha\text{-lactose.}$

Réactif B-per™ (thermo Fisher Scientific) pour lyse cellulaire.

1.4 Milieux de culture

Différents milieux de culture bactériens ont été utilisés dans ce travail et sont décrits cidessous :

Lysogeny broth (LB) (g/L) : 10g Tryptone, 5 g Extrait de levure, 5 g NaCl. Les bactéries transformées ont été sélectionnées sur milieu LB gélosé par ajout d'agar (20 g/L).

Auto-inducteur (AI) (1L) : ZY (g/L) 966,6 ml : 10g Tryptone, 5 g extrait de levure, MgSO4 2M 520 µl, 50x5052 20,8 ml, 20X NPS 51,6 ml. Ce milieu est utilisé pour permettre un rendement plus élevé de protéines exprimées.

Psi Broth (g/L) : 20g Tryptone, 5 g Extrait de levure, 5 g MgSO, pH 7.6. Milieu utilisé pour la préparation des bactéries compétentes.

Chapitre 2. Préparation des plasmides et bactéries

2.1 Clonages

Les protéines décrites dans ce travail, ont été exprimées dans *Escherichia coli* sans leur peptide de transit N-terminal. Les ADNc complets codant pour PII1 et SS4, clonés dans le plasmide pENT-D-Topo (Thermofisher, Rochester, USA) (Vandromme et al., 2019) ont été utilisés comme cibles pour la construction des vecteurs exprimant les différentes formes tronquées de SS4 et PII1. Les séquences peptidiques des constructions tronquées sont mentionnées dans l'Annexe 3.

Les fragments de l'ADNc de PII1 et SS4 ont été amplifiés avec les amorces décrites dans le **Tableau 1** à l'aide de la polymérase ultra-haute Kapa Hifi Hot start (Roche diagnostics, Autriche) et les produits de réaction en chaîne par polymérase (PCR) ont été transférés dans le plasmide pENT-D-Topo (Thermofisher, Rochester, USA). Ce vecteur d'entrée a été utilisé pour construire les vecteurs pET300-PII1, pML310-PII1 et pML333-PII1 en utilisant la technologie gateway dans le plasmide ChampionTM pET300, pML310 et pML333 (Thermofisher).

Les produits PCR correspondants aux fragments d'ADNc des constructions de SS4 tronquées SS4Δ349, SS4Δ403 et SS4Δ466 ont été amplifiés à l'aide de la polymérase Kapa Hifi Hot start et digérés par les enzymes de restriction BAMH1 et XHI1 avant d'être clonés dans les sites de restriction correspondants du vecteur pET-DUET-1 (Novagen) permettant l'expression de protéines marquées par une étiquette 6-His en N-terminal. Pour les expériences de co-expression, la séquence PII1 a été amplifiée en utilisant les amorces dans le **Tableau 1** permettant l'introduction des sites de restriction BamHI et Sall qui ont été utilisés pour cloner le produit PCR digéré dans le vecteur pET-DUET-1. Un deuxième produit PCR correspondant à la séquence d'ADNc SS4- Δ349 a été amplifié à l'aide des amorces cité dans le **Tableau 1** et le produit PCR a été cloné dans les sites de restriction EcoRV et Xhol du plasmide précédent, ce qui a permis la co-expression des protéines N-terminales 6-His-tag PII1 et SS4-Δ349. Les séquences peptidiques des constructions tronquées sont mentionnées dans l'Annexe 3.

Longueu Amorce anti-sens (kDa) Vecteur (aas) Tag Amorce sens de restriction 111 pET-300 6His-Nter 753 90 5'-CACCCTGCCATACTCGAATCTTGGC- 3' 5'-TTAGCGGTTAGTGAGCTCTGCAATACG-3' PII1-H2-H3 pET-300 461 53 6His-Nter Gateway PII1-∆71 pET-300 713 82 6His-Nter '-CACCAGTGCGAACGTGTTATTCGAT-3' '-TTACAGGATAACGCCTTCCATACA-3' Gateway PII1-H1-H2-H3 pET-300 494 57 6His-Nter 5'-CACCAGTGCGAACGTGTTATTCGAT-3' 5'-TTAGCGGTTAGTGAGCTCTGCAATACG-3' Gateway PII1-∆101 pET-300 682 78 6His-Nter 5'CACCCTGCCATACTCGAATCTTGGC-3' 5'-TTACAGGATAACGCCTTCCATACA-3' Gateway SS4-∆349 pET-Duet-1 661 75 6His-Nter 5'-CCCGGATCCCGAATGCACTGATCTATGG- 3' '-CCCTCGAGTTACGTGCGATTAGGAACAGCTC-3' BamHI et Xhol SS4-∆403 pET-Duet-1 637 69 6His-Nter 5'-CCCGGATCCCGAATCTTCAGAGAAAATTCAACAG-3" 5'-CCCTCGAGTTACGTGCGATTAGGAACAGCTC-3' BamHI et Xhol pET-Duet-1 SS4-∆466 574 63 6His-Nter -CCCGGATCCCCCGGTTGATGATATGCCTTGGGATTATTG-3 5'-CCCTCGAGTTACGTGCGATTAGGAACAGCTC-3' BamHI et Xhol PII1-SS4-A349 pET-Duet-1 780 90 6His-Nter 5'-GGGGATCCCGTGAATCATAAACAAAAAAGC-3' 5'-GGGGTCGACTTACAGGATAACGCCTTCCATAC-3 BamHI et Sall 661 5'-CCCGGATCCCGAATGCACTGATCTATGG-3' -CCCTCGAGTTACGTGCGATTAGGAACAGCTC-3 EcoRV et Xhol

Tableau 1. Amorces nucléotidiques utilisées pour la construction des protéines tronquées de PII1.

2.2 Préparation des bactéries chimio-compétentes

Les bactéries ont été incubées dans 2 ml dans le milieu LB à 37°C sous agitation pendant une nuit. À partir de la préculture, les bactéries ont été incubées dans 250 ml de milieu Psi Broth à 37°C jusqu'à atteindre 0,6 de DO₆₀₀. Ensuite, la culture bactérienne est placée 15 min dans la glace. Toutes les étapes suivantes sont effectuées à basse température, dans les conditions réfrigérées. Les bactéries sont récoltées par centrifugation 10 min à 4 000 rpm. Le culot est repris délicatement dans 100 ml de tampon TFBI froid puis incubé pendant 30 min dans la glace. Après nouvelle centrifugation, les bactéries sont re-suspendues dans 5 ml de tampon TFBII. La suspension bactérienne refroidie 20 min sur

glace puis aliquotée par fraction de 50 μ l. Les aliquots sont rapidement congelés dans l'azote liquide puis stockés à - 80°C.

2.3 Transformations bactériennes

Les cellules compétentes (50 µl) sont incubées avec 250 ng de la construction plasmidique pendant 30 min dans la glace. Les bactéries subissent ensuite un choc thermique 1 min à 42°C puis sont incubées 2 min dans la glace. Les cellules sont ensuite incubées avec 250 µl de milieu LB, placées 1 h à 37°C sous agitation puis étalées sur milieu sélectif LB gélosé contenant l'antibiotique de sélection pendant une nuit à 37°C.

Chapitre 3. Préparation des échantillons protéiques

3.1 Induction des gènes

Les protéines étudiées dans ce manuscrit sont connues pour être difficile à exprimer et à solubiliser. Pour réaliser les tests d'expression nous avons décidé de tester en parallèle de nombreuses conditions (**Tableau 2**).

Protocole standard utilisé pour les tests d'expression :

Tout d'abord nous avons testé en parallèle plusieurs souches bactériennes, connues pour optimiser les conditions d'expression et qui sont décrites dans le paragraphe dédié en commençant par des plasmides permettant l'introduction d'une étiquette 6-His à l'extrémité N-terminale de la protéine d'intérêt. Si ces constructions ne permettaient pas d'obtenir un échantillon soluble nous avons dans un second temps utilisé d'autres vecteurs (décrits au début du chapitre matériel et méthodes) permettant de fusionner les protéines d'intérêt avec des protéines (Thiorédoxine ou Maltose binding protéine) pour stabiliser l'échantillon). Dans ces conditions, une étiquette 6-His est également ajoutée pour permettre de standardiser le protocole de purification par chromatographie d'affinité sur métal immobilisé (IMAC). Pour chaque souche utilisée nous avons testé deux modes d'induction : l'induction par Isopropyl β -D-1-Thigalactopyranoside (IPTG) (à différentes températures, temps ou concentrations) ou auto-induction (AI).

L'effet de différents additifs ont également été testés dans les tampons de lyse cellulaire.

Les tests d'expression ont été réalisés avec de 50 ml de culture inoculés à partir d'une préculture réalisée sur la nuit à partir d'une colonie.

Pour l'induction par IPGT, la croissance bactérienne à 37°C est suivie en mesurant la turbidité de la suspension bactérienne à 600 nm. C'est à cette phase exponentielle de croissance (DO₆₀₀ nm) qu'on ajoute 0,5 mM d'IPTG dans le milieu de culture. L'ITPG est utilisé pour induire l'expression de protéines recombinantes placées sous le contrôle du promoteur lac. Il est utilisé comme un analogue de l'allolactose, un métabolite du lactose qui active la transcription de l'opéron lactose Lacl et en particulier du gène de la β -galactosidase, lacZ. Comme l'allolactose ou le lactose, l'IPTG se lie au répresseur de l'opéron Lacl ce qui empêche sa liaison à l'opérateur et induit donc la transcription des gènes puis la traduction des protéines.

L'induction par milieu auto-inductible permet quant à elle d'induire automatiquement l'expression des protéines recombinantes à partir du promoteur Lac après 3h de culture à 37°C sans ajouter d'inducteur artificiel tel que l'IPTG. Cette méthode est basée sur des composants du milieu qui sont métabolisés de manière différentielle. On utilise un milieu composé d'une solution ZYP (ZY, MgSO4 et NPS) qui apporte les nutriments nécessaires aux bactéries et d'une solution 5052 qui sert de source de carbone et contient du glucose, du lactose et du glycérol. Le glucose est métabolisé préférentiellement pendant la prolifération cellulaire, ce qui empêche l'absorption du lactose jusqu'à ce que le glucose soit épuisé, Lorsque le glucose est épuisé, le lactose est absorbé et lève la répression de l'opéron lactose en se liant au répresseur.

La culture est ensuite incubée à la température et la durée désirée selon le test d'induction.

Tableau 2. Récapitulatif des conditions d'expression testées pour chaque construction protéique. Les valeurs en gras représentent la meilleure condition d'expression pour chaque protéine.

Constructions protéiques	Vecteurs	Souches	Induction	Méthode de lyse	Tampons
Pil1	pET300	BL21	IPTG	Sonication	TrispH8
	pML375	C41	A	Emulsiflex	Hepes pH 7
	pML333	C43		B-Per	
	pML310	Rosetta Gami			
PI1-H2-H3	pET300	BL21	IPTG	Sonication	TrispH8
	pML333	Rosetta	A	Emulsiflex	Hepes pH 7
	pML310				Mes pH 6
					Bicine pH 9
					Naac pH 4
PII1-H1-H2-H3	pET300	BL21	IPTG	Sonication	Tris pH 8
			A	Emulsiflex	Hepes pH7
					. ispee prin
PII1-∆71	pET300	BL21	IPTG	Sonication	Tris pH 8
	pML333		A	Emulsiflex	Hepes pH 7
	pML310				
PII1-∆101	рЕТ300	BL21	IPTG	Sonication	Tris pH 8
			AI	Emulsiflex	Hepes pH 7
SS4∆349	pETDuet	BL21	IPTG	Sonication	Tris pH 8
		Rosetta	Al	Emulsiflex	Hepes pH 7
SS4∆403	pETDuet	BL21	IPTG	Sonication	Tris pH 8
		Rosetta	A	Emulsiflex	Hepes pH 7
SS4∆466	pETDuet	BL21	IPTG	Sonication	Tris pH 8
		Rosetta	AI	Emulsiflex	Hepes pH 7
PII1-SS4∆349	pETDuet	BL21	IPTG	Sonication	Tris pH8
			AI	Emulsiflex	Hepes pH 7

Afin d'affiner les conditions d'Al, nous avons réalisé un crible de 24 conditions d'induction en utilisant une concentration croissante de glucose (de 0 à 0,1%) et de lactose (de 0 à 0,3%). Cette méthode permet de contrôler manuellement le moment de l'induction et donc d'optimiser l'expression des protéines recombinantes. Cette expérience a été réalisée à 20°C, 30°C ou 37°C pendant 3h à partir d'une culture cellulaire de 2 ml.

Pour chaque essai, un volume de bactéries correspondant à 0,1 de DO à 600 nm est prélevé avant (NI) et après (I) induction. L'extrait est centrifugé à 6000 rpm et le culot bactérien est resuspendu dans 10 µl de bleu de charge pour SDS-PAGE et chauffé 5 min à 95°C. L'analyse de la fraction I permet de vérifier l'expression des protéines d'intérêt.
3.2 Lyses cellulaires

Après induction, les bactéries sont centrifugées 10 min à 6000 rpm et les membranes bactériennes sont lysées avec du B-per ou dans un tampon de lyse selon le volume de culture. Dans un petit volume de culture (50 ou 100 ml), les culots bactériens ont été lysés en ajoutant le réactif B-Per™ pendant 15 min à température ambiante pour l'expression de PII1 entière. Le B-Per™ est une solution à base de détergeant doux non ionique qui permet d'extraire les protéines solubles des cellules bactériennes sans les dénaturer.

Pour l'expression des constructions tronquées de PII1, de SS4 et des co-expressions, les culots ont été lysés par sonication. Dans ce cas, les culots bactériens sont repris préalablement dans 2 ml de tampon de lyse pour les petits volumes de culture et 20 ml pour les grands volumes de culture. La composition du tampon de lyse varie selon les tests d'expression et contient au minimum du NaCl 250 mM et une pastille d'inhibiteurs de protéases EDTA-free (Roche). Dans ce travail, plusieurs tampons ont été testés : Tris pH 8, Hepes pH 7, MES phH 6, Bicine pH 9 et acétate de sodium (NaOAc) pH 4.5 (voir **Tableau 2**). Le pH influence l'environnement de la protéine et donc sa solubilité.

Les culots bactériens ont été lysés par cycle de 3 x 15 sec avec 15 sec de repos entre chaque cycle, le tout effectué sur glace pour éviter la surchauffe des échantillons protéiques.

Dans un volume de culture supérieur à 250 ml, les culots bactériens ont été repris dans un tampon de lyse puis lysés par haute pression avec l'Emulsiflex C3 à 1500 bars (Avestin) ou la sonication 5 x 30 secondes. Les protéines solubles sont récupérées par centrifugation pendant 30 min à 8 000 rpm et la fraction soluble (S), préalablement additionnée avec 10µl de bleu de charge et chauffée à 95°C pendant 5 min est chargée sur SDS-PAGE pour vérifier la solubilité de la protéine.

Chapitre 4. Purification des échantillons protéiques

4.1 Purification par chromatographie d'affinité sur ions métalliques immobilisés (IMAC)

Les protéines ont été purifiées par IMAC grâce à l'étiquette 6-His incluse à l'extrémité Nterminale de chaque protéine. Les protéines solubles issues d'une culture de 50 ou 100 ml ont été incubées avec 50 µl résine Thermo Scientific™ HisPur™ Ni-NTA. Dans ce manuscrit on désignera la purification avec la résine de Nickel de purification par « Batch ». La résine est lavée dans un tampon de lyse avec ou sans imidazole 15 mM puis les protéines sont éluées par compétition avec 500 mM d'imidazole dans 50µl. Pour les volumes de culture supérieurs à 500 ml, les protéines sont incubées avec 500 µl de résine. Pour faciliter le protocole de purification nous avons également purifié les protéines sur colonne HisTrap™ de 5 ml avec l'ÄKTA go (ThermoFischer). Dans ce cas-là, la purification sera désignée par purification sur « colonne ». Dans les deux cas, Batch ou colonne, la résine est lavée dans un tampon de rinçage avec ou sans imidazole 15 mM pour se débarrasser des contaminants et protéines aspécifiques.

Enfin, les protéines d'intérêt sont éluées avec plusieurs volumes de colonne avec le tampon d'élution contenant 500 mM d'imidazole. La fraction éluée (E) ou non retenue (FT) sont additionnées avec 10µl de tampon de charge, chauffées 5 min à 95°C et déposées sur SDS-PAGE. L'analyse de la fraction E permet à la fois de vérifier la capacité de l'échantillon à se fixer à la résine IMAC et le faible volume de résine utilisé permet également de concentrer la protéine d'intérêt et de la visualiser si la quantité de protéines solubles est faible.

4.2 Purification par chromatographie d'exclusion stérique (SEC)

Pour améliorer la pureté de l'échantillon protéique, nous avons réalisé une deuxième étape de purification par chromatographie d'exclusion stérique (SEC) en utilisant la colonne

Superdex [™] 200 Increase 10/300 GL 25 ml (Cytiva) pré-équilibrée avec le tampon de lyse. La SEC permet de séparer les protéines en fonction de leur taille (dans ce cas les protéines allant de 10 à 600 kDa) grâce à la présence de billes poreuses qui retient les protéines les plus petites. Elle permet également d'éliminer la présence d'imidazole issue de l'étape de purification d'affinité. Nous avons injecté 500 µL d'échantillon protéique issue de la purification par affinité dans la colonne. L'élution des protéines est suivie par mesure de l'absorbance à 280 nm puis les protéines sont récupérées et les fractions protéiques purs sont concentrés à l'aide de tube de concentration Amicon 10K (Merck Millipore). La concentration finale des protéines est mesurée par l'absorbance à 280nm avec le spectrophotomètre NanoDrop™ (ND1000) de Thermo Scientific. Les échantillons protéiques, congelés dans l'azote liquide puis stockés à -80°C.

4.3 Renaturation des corps d'inclusion

Les corps d'inclusion ont été récupérés par centrifugation à 10 000 rpm 30 min à 4°C à partir d'une culture de 500 ml de PII1-pET300. Le culot cellulaire a été lavé 3 fois dans un tampon de lavage 1 :4 (p/v) Tris 50 mM, DTT 1mM, Tween20 20%, 2M urée. À ce stade les corps d'inclusion sont purs à 95%. Le mélange a été centrifugé à 10 000 rpm pendant 30 min à 4° C pour séparer les corps d'inclusion du surnageant. Les corps d'inclusion ont ensuite été dénaturés dans le tampon de dénaturation Tris 50 mM, DTT 1mM, 8 M d'urée. Les protéines solubilisées ont été repliées par l'ajout goutte à goutte du tampon de repliement Tris 50 mM pH 8, DTT 1 mM, imidazole 5 mM.

Chapitre 5. Identifications des protéines

5.1 SDS-PAGE

Avant le dépôt sur SDS-PAGE, les protéines ont été dénaturées dans un tampon de bleu de charge et incubées 5 min à 95 °C. Les différents échantillons (NI, I, S, FT et E) sont chargés sur les gels dénaturants de polyacrylamide à 10%. La migration est effectuée dans un système Mini Protean III, Bio Rad à 150 mV. Les protéines sont colorées à l'aide du bleu brillant de Coomassie G-25l (InstantBlue, expedeon, Blue Silver) après rinçage à l'eau distillée.

5.2 Western blot

Après SDS-PAGE, les protéines ont été transférées du gel vers une membrane de transfert PVDF Immobilon™ (Fisher scientific) par électrophorèse pendant 2h à 200 mV à 4°C. Après migration, la membrane est rincée à l'eau désionisée puis incubée 10 min avec la solution au rouge ponceau pour s'assurer du transfert correct des protéines. La membrane est rincée dans 15 ml de tampon TBST jusqu'à disparition du rouge ponceau puis placée dans 15 ml de la solution de saturation contenant 5% lait dans le tampon TBST pendant une nuit à 4°C. La membrane est lavée 3 fois dans le tampon TBST sous agitation puis incubée 1h à température ambiante et à l'obscurité dans la solution de saturation qui contient l'anticorps anti-6-His-HRP (peroxydase du raifort)(Thermofisher) dilué au 1/10 000[™]. Enfin, la membrane est lavée 3 fois dans 15 ml de tampon TBST puis les protéines sont révélées avec la solution de révélation du kit SuperSignal[™] West Pico (Pierce [™]).

5.3 Spectrométrie de masse

Les bandes des protéines PII1-H2H3 et PII1 et SSA349 issues du complexe PII1- SS4A349 pETDuet visibles sur SDS-PAGE ont été découpées du gel puis digérées par la trypsine et analysées par spectrométrie de masse par Désorption-Ionisation Laser Assistée par Matrice-temps de vol (MALDI-TOF) par la plateforme d'analyses Protéomiques et Peptides Modifiés (P3M) de l'Institut Pasteur de Lille ou par la plateforme MS for All Omics de l'UGSF.

Chapitre 6. Etude de l'activité de SS4

Des tests d'activité par zymogramme ont été réalisés pour étudier la capacité d'élongation de SS4. Dans ce travail nous avons utilisé les constructions tronquées SS4 Δ 403, SS4 Δ 466 et SS4 Δ 349 en utilisant les extraits bactériens solubles car pour cette expérience l'échantillon protéique n'a pas besoin d'être pur. Les échantillons protéiques sont incubés avec un tampon de charge sans SDS ni β -mercaptoéthanol pour ne pas les dénaturer puis déposés sur PAGE 10 % contenant 0,3 % de glycogène d'huître (Sigma). Le glycogène est structurellement assez similaire à l'amylopectine, bien qu'il soit plus fortement ramifié (tous les 10 résidus de glucose contre 30 pour l'amylopectine) et que les ramifications soient plus courtes. Ces chaînes pourront être allongées par SS4 et l'activité pourra être révélée à l'iode, formant des bandes brunâtres. L'iode ne se fixe qu'entre les chaînes de glucanes plus longues que celle du glycogène, qui lui n'est pas coloré à l'iode.

L'électrophorèse a été réalisée pendant 2h à 200 mA à 4°C à l'aide du système Mini Protean III de Bio-Rad. Le gel a ensuite été incubé pendant une nuit à température ambiante dans un mélange d'incubation de synthase contenant de l'ADPG qui le donneur de glucane. Les gels sont rincés trois fois dans l'eau désionisée afin d'éliminer les traces de β - mercaptoéthanol présentes dans le mélange d'incubation. L'activité d'élongation de SS4 a été révélée à l'iode. Des expériences de contrôle ont été réalisées en suivant la même procédure mais en omettant l'ADPG dans le mélange d'incubation. Cela permettra de vérifier que les bandes présentes dans le gel incubé avec ADPG sont dû à l'activité de SS4. En effet, SS4 utilise l' ADPG pour allonger les chaînes de glucanes du glycogène présent dans le gel. Son activité d'élongation sera ainsi visible. A l'inverse sans ADPG, SS4 n'est pas capable d'allonger les chaînes glycanniques.

Ensuite nous avons testé l'effet de PII1-H2H3 sur l'activité de SS4∆349 et SS4∆466. Les deux échantillons protéiques ont été incubés au moins 30 min ensemble avant d'ajouter le bleu de charge non dénaturant. Le protocole est ensuite le même que décrit précédemment. La fraction soluble de PII1-H2H3 a été également déposée pour vérifier que la protéine ne présente pas d'activité d'élongation. De plus, pour contrôler que l'effet de PII1-H2H3 sur l'activité de SS4∆349 n'est pas dû à l'encombrement stérique nous avons effectué un contrôle, dans les mêmes conditions qu'avec PII1-H2H3, en incubant l'échantillon soluble SS4∆349 avec la Sérum albumine bovine (BSA) en excès.

Pour compléter les contrôles, nous avons également déposé les extraits solubles des bactéries BL21(DE3) non transformées afin de différencier l'activité de SS4 et la présence éventuelle d'activité glycogène synthase (glgA) présente chez *E.coli*, qui est également capable d'allonger les chaînes glycanniques et donc susceptible d'avoir une activité visible sur gel après coloration à l'iode.

Enfin, pour tester non pas l'activité d'élongation mais plutôt l'activité d'initiation, nous avons réalisé un test d'activité sur gel natif qui ne contient pas de glycogène. Il a été montré que SS4 permet d'initier la synthèse de MOS à partir de l'ADPG. Pour cela l'extrait soluble de SS4 Δ 349 est déposé sur gel natif qui sera incubé dans le même mélange pour activité synthase avec ADPG. Pour servir d'amorce nous avons ajouté dans ce mélange, 0,3% de maltose et polyglucanes de DP 7-13. Le maltose a été décrit comme étant la plus petite amorce utilisable par les SSs et le polyglucane de DP 7-13 ressemble davantage au glycogène et servira de comparaison avec les tests sur gel contenant du glycogène.

78

Chapitre 7. Etude structurale et biophysique

7.1 Dichroïsme circulaire

La spectroscopie par dichroïsme circulaire (CD) permet d'étudier la structure secondaire (hélices α , feuillets β etc) des protéines en solution. La technique repose sur la capacité qu'ont les molécules optiquement actives d'absorber de façon inégale la lumière polarisée circulairement à droite et la lumière polarisée circulairement à gauche. De cette façon, on obtient un signal de dichroïsme circulaire à des longueurs d'onde qui correspondent à des bandes d'absorption de la molécule. L'analyse du spectre de dichroïsme circulaire d'une protéine dans la gamme de longueur d'onde autour de 240-320 nm permet de calculer la proportion d'hélices et de feuillets β . Les protéines entièrement composées d'hélices α présentent une bande intense négative à 222 nm et à 208 nm et une bande positive à 190 nm. Les protéines tout en feuillets β présentent une bande intense positive à 200 nm et une bande négative à 175 nm et 216 nm. Les protéines désordonnées présentent une bande positive à 220 nm et une bande négative à 180 nm et 200 nm. Il est également possible de prédire la présence de CC dans la structure si le ratio de 222/208 nm est supérieur à 1,00 (± 0.03) Les spectres obtenus sont déconvolués par les prédicteurs de structure tels que BestSel afin de calculer la proportion des structures secondaires de la protéine (Micsonai et al., 2018).

Les spectres de dichroïsme circulaire du rayonnement synchrotron (SR-CD) ont été mesurés sur la ligne de faisceau DISCO du synchrotron SOLEIL (Gif-sur-Yvette, France). Avant les mesures, le coefficient d'extinction elliptique molaire de l'Ammonium d- 10- Camphorsulfonate Ammonium (CSA) a été mesuré sur la ligne de faisceau et utilisé comme norme pour l'étalonnage de toutes les mesures de données (Miles et al., 2004). Les spectres CD ont été acquis à l'aide du logiciel IGOR (WaveMetrics) entre 170 à 291 nm. Les spectres des protéines et du tampon (20 mM Hepes/NaOH pH 7, 150 mM NaCl) ont été collectés consécutivement et représentent la moyenne de 3 mesures. La ligne de base du tampon a ensuite été soustraite des spectres et le traitement des données a été effectué à l'aide du logiciel CDToolX (Miles and Wallace, 2018).

L'échantillon protéique PII1-H2H3 concentré à 2,5 mg/ ml a été placé dans une cuvette en quarts de 50 nm. Les spectres CD ont été mesurés à 25°C.

Une autre expérience a été effectuée avec du 2,2,2-trifluoroéthanol (TFE) dans le but de tester l'influence du TFE sur la structure de PII1-H2H3. Le TFE permet de déstabiliser les CC en faveur d'hélices simples en créant un environnement hydrophobe favorable. Plusieurs tampons avec des concentrations croissantes en TFE (10, 20, 30, 50 ou 60% (v/v)) ont été incubés avec l'échantillon protéique pendant 10 min et les spectres CD ont été mesurés à 25°C. Nous avons veillé à conserver la même quantité de tous les échantillons en complétant le volume final avec le tampon (20 mM Hepes/NaOH pH 7, 150 mM NaCl). La ligne de base du tampon a ensuite été soustraite aux données mesurées et le traitement des données a été effectué à l'aide du logiciel CDToolX (Miles and Wallace, 2018).

7.2 Etudes bio-informatiques

2.2.1 Analyse des séquences peptidiques

La séquence peptidique de PII1 (Q8H1E5) a été analysée par le programme ProtParam du portail ExPASy (Gasteiger et al., 2005) pour analyser les caractéristique physicochimiques de la séquence.

7.2.2 Prédiction des structures secondaires

La séquence peptidique de PII1 (Q8H1E5) a été soumise aux méthodes PSIPRED (McGuffin et al., 2000) et Gor4 (Garnier et al., 1996) qui permettent de prédire la structure secondaire des protéines. PSIPRED (<u>http://bioinf.cs.ucl.ac.uk/psipred</u>) et Gor4 (<u>https://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html</u>) sont disponibles sous forme de serveur Web.

7.2.3 Modélisation structurale assistée par intelligence artificielle

Dans ce travail, nous avons utilisé plusieurs logiciels de modélisation : AlphaFold2 (Jumper et al., 2021), et MassiveFold (Brysbaert et al., 2024). Plus tard, avec l'arrivée d'AlphaFol3 (Abramson et al., 2024), les protéines et complexes ont été de nouveau modélisés.

AlphaFold est un logiciel d'intelligence artificiel développé par Deepmind qui prédit la structure 3D des protéines à partir de leur séquence en acide aminé. AlphaFold2 utilise les données de la Protein Data Bank (PDB), fournissant un modèle de prédiction pour des monomères, tandis qu'AlphaFold3 est une version améliorée qui permet de générer cinq modèles de prédiction. Ce dernier à l'avantage de proposer une précision plus élevée au niveau de la structure des protéines et de la structure des interactions protéine-protéine(Abramson et al., 2024).

MassivFold est une version optimisée et personnalisable d'AlphaFold qui permet d'exécuter plusieurs prédictions en parallèle en réduisant le temps de prédiction de plusieurs mois à quelques heures. Il permet d'accéder à une plus grande diversité structurelle guidé par un classement de confiance des modèles et permet d'améliorer les capacités de modélisation des structures monomériques et des complexes protéiques. Il permet également de réaliser plus d'étape de « recyclage » des modèles avec un seuil de tolérance afin d'avoir accès à une plus grande variabilité structurale. En dessous d'un certain seuil, si la distance entre la structure actuelle et précédente est trop grande, le recyclage s'arrête. Ainsi il peut prédire cinq modèles pour les monomères et quinze pour les multimères. (Pour aller plus loin voir : https://www.researchsquare.com/article/rs-4319486/v1).

Pour chaque modèle, nous avons utilisé des critères de sélection pour qualifier les modèles prédits. Dans le cas de la prédiction des monomères et multimères, trois paramètres permettent de connaitre la précision et la fiabilité du modèle :

- pLDDT (predicted Local Distance Difference Test) qui compare localement la structure expérimentale et la prédiction. Il donne une mesure de la qualité de la

81

prédiction pour chaque acide aminé. Il est donné sous forme d'un score entre 0 et 100 où 100 signifie que le modèle est très fiable. pLDDT > 90 : régions modélisées avec une grande précision pLDDT entre 70 et 90 : régions modélisées avec une bonne précision pLDDT entre 50 et 70 : régions prédites avec une faible précision pLDDT inférieur à 50 : régions prédites avec une très faible précision ou régions désordonnées.

- PAE (Predicted Alignment Error) indique, pour chaque position X, la différence entre la structure expérimentale et la structure prédite lorsque les deux structures sont alignées à la position Y. C'est donc une mesure de la qualité de la prédiction de la position des domaines les uns par rapport aux autres.
- pTM (predicted Template Modeling score) mesure la différence entre la structure expérimentale et la structure prédite, normalisée par la longueur de la protéine. Il varie de 0 à 1 où 1 est une correspondance parfaite.

Dans le cas des multimères uniquement, un autre paramètre rentre en compte pour la précision du modèle : ipTM (interface predicted Template Modeling score) mesure la qualité de la prédiction de l'interface protéine-protéine en utilisant le même score que le score pTM.

Les modèles moléculaires avec les meilleures valeurs de pLDDT et PAE, pTM ou ipTM ont été utilisés dans ce travail. Malgré tous les progrès considérables en matière de modélisation, il existe encore des limitations en ce qui concerne la modélisation des atomes (en conflit), voire des chaînes, qui se chevauchent (Abramson et al., 2024). Certains de nos modèles, avec des scores pTM ou iPTM (dans le cas des complexes) trop faibles (score inférieur à 0,4) n'ont pas été montrés dans ce travail.

Les structures et les surfaces ont été visualisées à l'aide de PyMOL (Schrödinger and Delano., 2020). Certains modèles ont été alignés en utilisant PyMOl qui permet de calculer le RMSD entre deux structures (root mean square deviation). Plus le RMSD est proche de 0, plus l'alignement entre deux protéines est important.

7.2.4 Alignements multiples des séquences

Les séquences des orthologues de *At*PII1 ont été récupérés depuis le site TAIR (https://www.arabidopsis.org/), banque de données du génome d'*Arabidopsis thaliana* et qui fournit des informations telles que les orthologues, l'expression des gènes, les stocks d'ADN etc... Les séquences de 27 orthologues ont été alignées avec *At*PII1 en utilisant BLASTp (Altschul et al., 1990) et l'alignement multiple a été calculé avec COBALT (Papadopoulos and Agarwala., 2007) avec les paramètres par défauts.

7.2.5 Prédiction coiled-coils

La séquence peptidique de PII1(Q8H1E5) a été soumise aux programmes MARCOIL et DeepCoil2 et PCOILS afin de prédire les structures en CC (Delorenzi and Speed., 2002; Gruber et al., 2006; Ludwiczak et al., 2019). Ces méthodes sont disponibles sous des versions web.

7.2.6 Prédiction des résidus conservés

Les résidus conservés ont été identifiés par deux méthodes. Dans la première méthode, les résidus conservés ont été identifiés à partir de l'alignement entre les orthologues de *At*PII1 avec Blasp et Cobalt (Papadopoulos and Agarwala., 2007). L'alignement final a été soumis à EndScript qui permet de mettre en évidence les résidus conservés et est disponible sur le serveur : <u>https://endscript.ibcp.fr/ESPript/ENDscript/</u> (Robert and Gouet., 2014). Dans la deuxième méthode, la structure AF3 de PII1 a été soumise à ConSurf disponible sur le serveur : <u>https://consurf.tau.ac.il/consurf_index.php</u> (Ashkenazy et al., 2016). La protéine est d'abord soumise à Blast à partir de la base de données protéiques UniRef. Ensuite, ConSurf recherche des orthologues et fait l'alignement avec MAFFT et l'alignement de séquence multiple. Enfin, les résidus conservés sont colorés en fonction de leur niveau de conservation : du cyan (variable) à rouge (conservé).



Ce projet de thèse a pour but de caractériser la fonction de deux protéines du métabolisme de l'amidon impliquées dans l'étape d'initiation, PII1 et SS4. Mes résultats se structurent en cinq chapitres.

Les trois premiers chapitres concernent la production des protéines PII1, SS4, ainsi que du complexe PII1/SS4. Le but initial de ce projet était l'étude structurale de PII1 et de son interaction avec SS4. Du fait de la grande difficulté que j'ai rencontrée pour obtenir des échantillons protéiques solubles dans un premier temps et monodisperses et stables dans un deuxième temps, une grande partie de mon travail a consisté à obtenir des échantillons compatibles avec des études structurales. De nombreux tests d'expression et constructions ont dû être réalisés pour obtenir des échantillons protéiques utilisables et ont constitué plus de la première moitié de mon travail de thèse. L'ensemble des conditions testées est listé par protéine dans les **Tableaux 4 à 12** à la fin du chapitre 3.

Dans un second temps, du fait de la propension des deux protéines à former des agrégats, nous avons préféré les étudier par des approches de bio-informatique combinées à des approches biochimiques et biophysiques qui seront décrites dans le chapitre 4.

Enfin, dans la cinquième partie, nous avons réalisé les tests d'activité enzymatique à partir des constructions tronquées de SS4. Puis de façon à avancer dans la caractérisation de la fonction de PII1, nous avons étudié l'effet de la présence de PII1 sur l'activité de SS4.

L'ensemble de ces travaux ont fait l'objet d'un article de recherche actuellement soumis au journal Plant Physiology, en cours de révision (**Annexe 1**).

Une dernière partie de mes travaux concerne les protéines de l'assemblage et de la protection du grain : LESV et ESV1 qui constituaient le sujet de thèse de Rayan Osman dans notre équipe. J'ai participé à ce travail notamment dans les mesures et traitements de données de SAXS, CD et de microscopie de fluorescence mesurés sur les lignes SWING, DISCO-CD et DISCO-imaging du synchrotron SOLEIL. A ce titre, je suis deuxième auteure sur un article publié dans Plant Physiology cette année (Annexe 2). Après le départ de Rayan Osman, j'ai finalisé un projet visant à mettre en évidence le rôle de ESV1 dans la protection des grains d'amidon qui fera l'objet d'une publication avec d'autres données en cours d'écriture et sur laquelle Rayan Osman et moi serons co-premières auteures.

Chapitre 1. Analyse de l'expression de PII1

Dans ce chapitre j'ai entrepris la production de la protéine PII1 sous sa forme entière puis tronquée pour faciliter sa solubilisation. Un travail important d'optimisation des conditions de production des échantillons protéiques a été mené en modifiant les conditions d'expression, la souche bactérienne ou les vecteurs d'expression. S'ajoute à cela l'optimisation de la purification en utilisant diverses approches de séparation des protéines. Ce travail d'expression et de purification a pour but de préparer des échantillons de PII1 pour son étude par des approches de biophysique ou de biologie structurale.

1.1 Optimisation des conditions d'expression et de la solubilité de PII1

1.1.1 Tests d'expression

À mon arrivée au laboratoire, la protéine PII1 venait d'être caractérisée. Dans ce contexte, des premiers essais de production de PII1 recombinante chez la bactérie *E.coli* BL21(DE3) semblaient indiquer qu'une fraction de la protéine exprimée était soluble. La construction utilisée a été synthétisée avec des codons optimisés pour son expression par *E. coli* et insérée dans un plasmide pENTR-D-TOPO puis dans le vecteur d'expression pET300 permettant l'insertion d'une étiquette 6-His à l'extrémité N-terminale de la protéine. Les résultats préliminaires obtenus avaient montré une expression visible sur SDS-PAGE dans la fraction induite mais peu dans la fraction soluble (résultats non publiés). Répéter ce résultat a constitué le point de départ de mon travail.

PII1 clonée dans le vecteur pET300 a été produite dans la souche BL21(DE3) en testant deux méthodes d'induction, soit par l'ajout d'IPTG 0,1 mM soit par AI. Les cultures ont été incubées toute une nuit à 20°C. Les protéines solubles ont ensuite été récupérées après lyse des cellules bactériennes puis purifiées sur résine de nickel par batch. Les résultats sont présentés dans la **Figure 22**.



Figure 22. Expression et solubilité de PII1. SDS-PAGE 10% coloré au bleu de Coomassie de PII1-pET300 exprimée dans deux conditions d'induction différentes (0,1 mM d'IPTG ou auto-induction AI). NI = fraction non induite. I = fraction induite. S = soluble. E = fraction purifiée sur résine de Nickel. La flèche correspond à la région de migration attendue pour PII1.

L'analyse des résultats montre que PII1 est exprimée en quantité similaire quelles que soient les conditions d'induction choisies (Pistes I). En revanche, dans les deux cas on observe très peu de protéines solubles (Pistes S) et aucune bande correspondant à PII1 visible après purification (Pistes E).

Des approches de génomique fonctionnelle ont montré que le niveau d'expression et de solubilité d'une protéine recombinante peut varier de façon drastique en fonction de la souche bactérienne utilisée, de ce fait nous avons testé l'expression de PII1 dans d'autres souches d'expression (Jonasson et al., 2002). Certaines d'entre elles comme Rosetta-Gami sont des souches dérivées de BL21(DE3) qui améliorent l'expression des protéines car contiennent des codons rarement utilisés chez *E. coli*. D'autres souches comme C41 ou C43 permettent de conférer une résistance en cas de protéines toxiques. La protéine a été exprimée dans les deux conditions d'induction, à savoir avec l'induction des gènes par l'IPTG ou Al **(Figure 23).**

Afin de tester tous ces paramètres, j'ai mis en place un protocole standard (décrit dans la partie Matériel et Méthodes). Brièvement, nous avons utilisé 50 ml de culture bactérienne puis induit l'expression des protéines. Les fractions protéiques avant (NI) et après (I) induction, soluble après lyse (S) ont été analysées sur SDS-PAGE. Les résultats sont présentés **Figure 23.** Tous les tests d'expression réalisés avec PII1 sont listés dans le **Tableau 4.**



Figure 23. Optimisation de l'expression de PII1 en fonction de la souche et du vecteur d'expression. (A) 10% SDS-PAGE coloré au bleu de Coomassie de PII1-pET300 exprimée dans les souches C41, C43 ou Rosetta-Gami. (B) 10% SDS-PAGE coloré au bleu de Coomassie de PII1 exprimée dans BL21(DE3) et clonée dans le vecteur pML310 qui contient une étiquette 6- His et une fusion avec la Thioredoxine (TRX) et dans pML333 qui contient une étiquette 6-His et une fusion gProtein (MBP). NI=fraction non induite. I =fraction induite. S = fraction soluble. La flèche correspond à la région de migration attendue pour chaque construction.

L'analyse des résultats montre que la quantité de protéines exprimées est plus faible dans les souches C41, C43 et Rosetta-Gami que dans la souche BL21(DE3) et que celles-ci n'ont pas permis d'améliorer la quantité de protéines solubles. Nous avons, en conséquence, décidé de continuer les tests d'expression dans la souche BL21(DE3) pour caractériser le vecteur d'expression le plus adapté pour obtenir de la protéine soluble (**Figure 23**). En effet, fusionner la protéine à une protéine chaperonne ou un peptide soluble peut permettre d'améliorer sa solubilité. Pour cela PII1 a été transférée dans le plasmide pML333 ou pML310 par LR réaction à partir du pENTR-D-TOPO-PII1. pML333 permet d'ajouter une étiquette 6-His pour faciliter la purification et une protéine de liaison au maltose (MBP) qui facilite le repliement protéique. pML310 permet d'ajouter la protéine thioredoxine (TRX) qui peut également faciliter la solubilisation (Figure 23).

Les résultats ne montrent aucune amélioration de la solubilité par rapport à l'expression dans le vecteur pET300. En résumé, le meilleur résultat que nous ayons obtenu pour l'expression de PII1 est dans la souche BL21(DE3) lorsque PII1 est clonée dans le vecteur pET300. Les conditions d'expression qui permettent d'obtenir une quantité de PII1 détectable sont réalisées dans le milieu AI à 20°C pendant une nuit. C'est donc dans ces conditions que nous avons exprimé PII1 pour la suite des expériences.

Afin d'améliorer la solubilité de PII1 j'ai décidé de tester différents milieux d'AI (**Figure 24**). Pour cela j'ai fait varier la concentration des trois sources de carbone présentes dans le milieu ZY-50x5052 qui compose le milieu d'AI à savoir le glucose, le glycérol ou le lactose. Les protéines ont été exprimées dans 24 conditions différentes. Un gradient de concentration a été appliqué pour le glucose (0 à 0,1%), le lactose (0 à 0,3%) et le glycérol (0 à 0,3%) en utilisant une base constante du milieu ZYP comme décrit dans la section du matériel et méthodes **(Tableau 3).** De plus, la production des protéines a été réalisée à trois températures d'incubation différentes : à 20°C, 30°C et 37°C pendant 3h.

Tableau 3. Compositions des milieux auto-inductibles utilisés pour l'optimisation de l'expression et de la solubilité de PII1-pET300 dans BL21(DE3).

	A1	A2	A3	A4	A5	A6	B1	B2	B3	B4	B5	B6
%glucose	0	0	0	0	0	0	0.05	0.05	0.05	0.05	0.05	0.05
%lactose	0	0.0375	0.075	0.15	0.225	0.3	0	0.0375	0.075	0.15	0.225	0.3
%glycérol	0.3	0.225	0.15	0.075	0.0375	0	0.3	0.225	0.15	0.075	0.0375	0
	C1	C2	C3	C4	C5	C6	D1	D2	D3	D4	D5	D6
%glucose	C1 0.075	C2 0.075	C3 0.075	C4 0.075	cs 0.075	C6 0.075	D1 0.1	D2 0.1	D3 0.1	D4 0.1	D5 0.1	D6 0.1
%glucose %lactose	C1 0.075 0	C2 0.075 0.0375	C3 0.075 0.075	C4 0.075 0.15	C5 0.075 0.225	C6 0.075 0.3	D1 0.1 0	D2 0.1 0.0375	D3 0.1 0.075	0.1 0.15	D5 0.1 0.225	D6 0.1 0.3
%glucose %lactose %glycérol	C1 0.075 0 0.3	C2 0.075 0.0375 0.225	C3 0.075 0.075 0.15	C4 0.075 0.15 0.075	C5 0.075 0.225 0.0375	C6 0.075 0.3 0	D1 0.1 0.3	D2 0.1 0.0375 0.225	D3 0.1 0.075 0.15	D4 0.1 0.15 0.075	0.1 0.225 0.0375	0.1 0.3 0



Figure 24. Optimisation de l'expression de PII1 par auto-induction. SDS-PAGE 10% coloré au bleu de Coomassie de PII1-pET300 induite dans les conditions d'auto-induction du Tableau 3, 3h à 20,30 ou 37°C. I = fractions après induction et S = fractions solubles. glyOH = glycérol, Glc = Glucose. La flèche correspond à la zone de migration attendue pour PII1.

Le résultat du criblage à 20°C montre que PII1 est faiblement exprimée quel que soit le milieu d'AI. Il est probable que cette température soit trop basse pour voir une expression protéique après seulement 3h d'induction. En revanche, l'expression de PII1 est plus élevée à 30°C avec une forte expression de PII1 dans presque toutes les conditions d'AI. Cependant la quantité de protéines solubles est la même dans chaque condition à un taux relativement bas. Enfin à 37°C, l'expression de PII1 est variable selon les milieux utilisés mais inférieure à l'induction à 30°C. Bien qu'il y ait des variations dans les différents essais, les quantités de protéines solubles sont trop faibles pour les expériences de biologie structurale que nous souhaitons réaliser par la suite.

Après ces nombreux tests, nous avons décidé d'utiliser la condition d'expression la plus prometteuse en termes de quantité de protéines solubles, à savoir dans le vecteur pET300 et la souche BL21(DE3), induite par auto-induction à 20°C toute la nuit. C'est ce mode d'expression qui a été appliqué par la suite dans ce travail.

1.1.2 Expression de PII1-pET300

Pour récupérer un maximum de matériel protéique soluble nous avons continué les essais de production de PII1 en augmentant volume de culture. Pour cela nous avons appliqué le protocole décrit précédemment sur un volume de culture à 500 ml. Le culot bactérien est resuspendu dans un tampon de lyse 50 mM Tris pH 8, NaCl 300 mM. La lyse chimique étant plutôt réservée pour de petits volumes de culture, les bactéries sont lysées par pression lorsqu'il s'agit de volumes de culture plus importants. Les fractions sont ensuite analysées par SDS-PAGE et la présence de PII1 a été vérifiée par spectrométrie de masse MS-MS MALDI-TOF (**Figure 25**).



Figure 25. Expression et caractérisation de PII1-pET300 produite dans BL21(DE3). (A) SDS-PAGE 10 % coloré au bleu de Coomassie de PII1-pET300 induite par auto-induction toute une nuit à 20°C. NI =fraction non induite. I = fraction induite. S = fraction soluble. C = fraction non soluble. (B) Analyse par spectrométrie de masse MS-MS de la fraction soluble issue de (A). La flèche correspond à la zone de migration attendue pour PII1.

Les résultats obtenus montrent que PII1 est exprimée (Piste 1) et soluble en faible quantité (Piste S). La bande correspondante à PII1 de la piste soluble a été traitée à la trypsine et les peptides obtenus ont été analysés par spectrométrie de masse (encadré rouge). PII1 a été identifiée en rang 1 avec 81% de recouvrement de séquence, ce qui confirme la présence de PII1 dans la fraction soluble. En revanche, on remarque que la majeure partie des protéines sont concentrées dans la fraction insoluble du culot bactérien (Piste C). A ce stade nous pouvons considérer PII1 comme étant une protéine majoritairement insoluble et qui présente une quantité très limitée de protéine solubles.

Pour améliorer le rendement de l'obtention de protéines solubles, j'ai testé l'effet de différents tampons de lyse additionnés d'additifs et/ou détergents. J'ai ajouté 0,1% de Triton X100, un détergent doux afin de masquer d'éventuelles régions hydrophobes pouvant empêcher la solubilisation des protéines. J'ai également ajouté de l'EDTA à 15 mM ou 1 mM pour prévenir de la dégradation des protéases et également testé l'effet de la concentration en sels sur la solubilité en rajoutant 1M ou 2M de NaCl. Enfin j'ai réalisé un autre tampon de lyse avec un mélange de NaCl 1M, EDTA 1mM et/ou Triton X100 0,1 % et du DTT 0,1mM un

réducteur qui facilite la solubilisation et protège de l'oxydation des cystéines. Les résultats obtenus sont présentés en **Figure 26**.



Figure 26. Optimisation de la solubilité de PII1 par l'ajout d'additifs dans le tampon de lyse. (A) SDS-PAGE 10% de PII1-pET300 solubilisé dans des tampons de lyse additionnés de différents additifs sur la base de 50 mM de Tris pH8. (*B*) SDS-PAGE 10% de PII1-pET300 solubilisé dans un tampon contenant du Tris pH 8, DTT 0,1mM, EDTA 1 mM et 5% de glycérol. NI = fraction non induite, I= fraction induite par auto-induction à 20°C 12h, S= fraction soluble. La flèche correspond à la zone de migration attendue pour PII1.

L'analyse des résultats montre que PII1 est bien exprimée (Piste I), par contre dans les pistes solubles (Piste S), PII1 est très faiblement présente ce qui signifie que l'ajout d'additifs ou détergents n'améliore pas la quantité de protéines solubles. A la suite de ces résultats et pour donner une chance de solubilisation, j'ai entrepris de solubiliser PII1 à partir des corps d'inclusion, fraction dans laquelle la protéine est en quantité très importante.

Les corps d'inclusion ont été préparés en les lavant plusieurs fois dans un tampon avec 2M d'urée et 20% de Tween20 (*cf.* protocole entier dans la section Matériel et Méthodes). L'urée permet de détacher les protéines des membranes bactériennes et le Tween 20 est un détergent non ionique qui permet de solubiliser la protéine. La protéine PII1 a été dénaturée avec 8M d'urée puis progressivement repliée dans un tampon de renaturation sans urée. Lors de cette étape, un précipité blanc est apparu dans la solution de renaturation indiquant que PII1 ne s'est pas repliée correctement ou n'est pas soluble dans le tampon (résultats non représentés). En conclusion, bien qu'une quantité relativement faible de PII1 soit soluble, la majorité des protéines se trouvent dans les corps d'inclusion. Cette quantité de protéine est insuffisante et incompatible avec l'idée de pouvoir continuer son analyse structurale.

Je me suis alors interrogée sur les raisons responsables de cette insolubilité. J'ai décidé d'orienter mes recherches vers deux possibilités pour élaborer la suite de mon travail de recherche.

- Une ou plusieurs régions de PII1 pourraient être responsables de l'insolubilité et devront, si possible, être identifiées. Des constructions tronquées de ces régions pourraient être plus solubles.

- PII1 requiert la présence d'un ou plusieurs de ses partenaires pour être soluble.

1.2 Réalisation et expression des constructions tronquées de PII1

1.2.1 Clonage des constructions tronquées

Nous avons émis précédemment l'hypothèse de la présence de régions pouvant gêner la solubilisation de PII1. Pour les identifier, nous nous sommes basés sur le modèle de PII1 tel qu'il est prédit par AF2, disponible sur <u>https://alphafold.ebi.ac.uk/</u> à partir du code Uniprot Q8H1E5 (**Figure 27**) (Jumper et al., 2021).



Figure 27. Analyse structurale de PII1 et conception des constructions tronquées (A) Modèle AlphaFold2 de PII1 représenté en cartoon dans la surface moléculaire. Les six hélices sont numérotées de H1 à H6 de l'extrémité N-terminale à C-terminale de la protéine. Les régions sont colorées en fonction de l'indice de confiance du modèle (pLDDT) indiqué à droite. (B) Représentation schématique des constructions de PII1. Δ = Délétion de X acides aminés. H=hélices.

La structure de PII1 est composée de 6 hélices (H1 à H6) de taille variable (**Figure 27**). Les hélices H2 (104 à 312 résidus) et H3 (320 à 563 résidus) sont longues et susceptibles de pouvoir s'associer entre elles pour former les régions CC prédites. A ce stade de ma thèse ce modèle AF2 ne prédit pas de régions CC, nous verrons par la suite que nous avons réussi à améliorer ce modèle par d'autres approches mais c'est avec le modèle décrit ici que j'ai défini les bornes des constructions tronquées.

Les hélices H1 (72 à 90 aa), H4 (599 à 679 aa), H5 (683 à 725 aa) et H6 (743 à 768 aa) sont plus courtes mais également prédites avec un haut degré de confiance. Le modèle obtenu par AF2 prédit une structure en hélice avec un score d'indice de confiance élevé à très élevé sur l'ensemble de la structure excepté au N-terminal de H1 et entre H3 et H4 où l'on observe une région désordonnée (pLDDT < 50). Nous avons estimé que ces régions désordonnées pourraient être à l'origine de l'insolubilité de PII1 sous sa forme entière.

Pour augmenter la solubilité de PII1 j'ai réalisé quatre constructions tronquées en supprimant la région N-terminale désordonnée (**Figure 27B**). Ensuite j'ai gardé ou pas l'hélice 1 (H1) qui semble être prédite avec un indice de confiance élevé. Les hélices H2 et H3 ont été conservées dans toutes les constructions tronquées car elles sont susceptibles de pouvoir former des CC, comme prédit dans les études antérieures (Vandromme et al., 2019) mais aussi parce que ces deux hélices sont conservées parmi les orthologues PII1. L'analyse des régions conservées sera discutée dans le paragraphe 4.1.6 du chapitre 4. Enfin j'ai gardé ou non l'extrémité C-terminale qui comprend H4, H5 et H6 pour éliminer la région désordonnée entre H3 et H4 qui peut gêner la solubilité de PII1 et dont la structure « très rectiligne » n'est probablement pas prédite correctement dans le modèle.

En résumé, à partir du modèle AF2, nous avons réalisé quatre constructions tronquées de PII1 (Q8H1E5) clonées dans le vecteur pET300 (**Figure 27B**). La première construction, PII1Δ71, comporte la séquence entière sans les 71 premiers résidus qui sont prédits désordonnés. De même pour la deuxième construction, PII1Δ101, qui exclut en plus la région H1. La troisième construction PII1-H2H3 contient seulement les H2 et H3. Enfin PII1-H1H2H3 contient uniquement les hélices H1, H2 et H3.

1.2.2 Tests d'expression

Les tests d'expression pour ces constructions ont été réalisés avec le même protocole que celui décrit dans le matériel et méthodes et nous avons également comparé le mode d'induction par IPTG 0,5 mM ou par auto-induction à 20°C toute la nuit (**Figure 28**). Le PM attendu des constructions tronquées est de 82 kDa pour PII1Δ71, 78 kDa pour PII1Δ101, 57 kDa pour PII1-H1H2H3 et 53 kDa pour PII1-H2H3.



Figure 28. Expression des constructions tronquées de PII1. SDS-PAGE 10% coloré au bleu de Coomassie des différentes constructions de PII1 .NI = fraction induite. I =fraction induite. S = fraction soluble. E = fraction purifiée sur résine de Nickel. La nature des constructions et le mode d'induction sont indiqués respectivement en haut et en bas de chaque gel. Les flèches correspondent à la zone de migration attendue pour les constructions de PII1.

Les résultats montrent que toutes les constructions protéiques sont exprimées quelles que soient les conditions d'induction mais on observe des différences d'expression et de solubilité. Ainsi, les constructions PII1Δ71 et PII1Δ101 sont les plus exprimées mais sont peu solubles (Piste S). En revanche, les constructions PII1-H1H2H3 et PII1-H2H3 permettent l'obtention de protéines solubles en grande quantité4. La suppression des 70 premiers résidus et la région prédite désordonnée entre H3 et H4 semblent avoir facilité la solubilisation de PII1.

PII1-H1H2H3 est plus soluble après induction par AI tandis que PII1-H2H3 est autant soluble après induction avec l'IPTG que par auto-induction. On observe cependant que PII1-H2H3 permet l'obtention d'une plus grande quantité de protéines par rapport aux autres constructions protéiques. L'ensemble des travaux d'optimisation, d'expression et de purification effectué sur toutes les constructions durant ce travail de thèse est listé dans les **Tableaux 5 et 6**. J'ai choisi de travailler avec PII1-H2H3 pour la suite de ce travail.

1.2.3 Purification de PII1-H2H3

Afin de collecter plus de matériel protéique, j'ai exprimé PII1-H2H3 dans un plus grand volume de culture (500 ml). Les cultures ont été réalisées dans les conditions décrites dans le chapitre précédent. Après solubilisation, le protocole de purification de la protéine a été défini après l'utilisation de plusieurs approches de purification par chromatographie : une étape de chromatographie d'affinité suivie d'une étape de chromatographie d'exclusion stérique.

Première étape : Chromatographie d'affinité sur ions métalliques immobilisés

Lors des tests d'expression, étant donné que nous travaillions avec des petits volumes de culture, j'ai purifié les protéines en tube sur un petit volume résine de Nickel. Afin de faciliter le protocole de purification et de pouvoir suivre l'élution des protéines, j'ai utilisé ici une colonne d'affinité IMAC (Histrap excel de 5 ou 1 ml, Cytiva) et un système de purification AKTA (Cytiva) qui offre l'avantage de pouvoir travailler de façon reproductible et de fractionner l'élution des protéines purifiées (**Figure 29**).



Figure 29. Purification par chromatographie d'affinité de PII1-H2H3-pET300 sur colonne IMAC HisTrap Excel 5 ml. A gauche chromatogramme indiquant le profil d'élution à droite le SDS-PAGE analysant l'induction et la purification de l'échantillon. NI = fraction non induite. I = fraction induite. S = fraction soluble. FT = fraction non retenue. Les pistes 1 et 2 correspondent aux fractions après purification colonne de Nickel. La flèche correspond à la zone de migration attendue pour PII1-H2H3.

L'analyse des résultats montre une quantité importante de protéines exprimées et solubles (Piste I et S). La majeure partie de PII1-H2H3 n'est cependant pas retenue sur la résine (piste FT), ce qui peut s'expliquer soit par une saturation de la colonne du fait de la grande quantité de protéines présentes, soit par l'inaccessibilité de l'étiquette 6-His pour une partie des protéines. Compte tenu de la structure prédite en CC de PII1-H2H3, il est probable qu'elle s'auto-agrège et qu'une partie des étiquettes soient masquées dans les complexes. Cependant, contrairement à la forme entière de PII1, une partie des protéines solubles est retenue sur la colonne puis éluée. On note la présence de contaminants à l'issue de cette étape que nous avons voulu éliminer avec une deuxième étape de purification.

Deuxième étape : chromatographie d'exclusion stérique

Afin d'éliminer les contaminants encore présents à l'issue de la première étape et pour évaluer la monodispersité de la solution de PII1-H2H3, nous avons effectué une deuxième étape de purification par chromatographie d'exclusion stérique avec une colonne Superdex 200 10/300 GL (**Figure 30**).



Figure 30. Purification par chromatographie d'exclusion stérique de PII1-H2H3-pET300 sur colonne Superdex 200 10/300 GL. A gauche le profil d'élution indique les volumes d'élution de chaque pic. Les étoiles indiquent les fractions analysées par SDS-PAGE (à droite). La flèche correspond à la zone de migration attendue pour PII1-H2H3.

Sur le chromatogramme, on constate la présence d'au moins six pics d'intensité variable dont j'ai analysé le contenu par SDS-PAGE. Le pic 4 contient plusieurs pics non séparés. La protéine PII1-H2H3 est éluée majoritairement dans le pic 3 (12 ml). En revanche, PII1 est présente dans tous les autres pics, notamment le pic 1, élué très tôt. Cela peut être expliqué en partie par la structure non globulaire de PII1-H2H3 qui pourrait alors passer entre les billes de silices (ou ne pas en ressortir facilement). Lors des nombreuses purifications, j'ai fréquemment observé que PII1-H2H3 était éluée très tôt (entre 7 et 9 ml), probablement à cause du fait qu'elle se regroupe en formant une grosse molécule sous la forme d'agrégats solubles qui passent hors des billes de silice. Du fait de la structure de la protéine, il est probable que les volumes d'élution ne soient pas informatifs ni de la polydispersité de la solution ni du PM, ni de l'état d'oligomérisation.

Pour optimiser le profil d'élution et la qualité de l'échantillon, j'ai testé d'autres tampons, avec différents pH, pour équilibrer la colonne de chromatographie, à savoir, Tris pH8, Bicine pH 9, Mes pH 6, Hepes pH 7, et acétate de sodium pH 4,5 (**Figure 31**). Les fractions ont été analysées sur SDS-PAGE 10%.



Figure 31. Optimisation du choix du tampon d'élution pour la purification de PII1-H2H3 par chromatographie d'exclusion stérique sur Colonne Superdex 200 10/300 GL. (A) Profils d'élutions de PII1-H2H3 dans le tampon Tris pH 8 (bleu), Mes pH 6 (orange) ; acétate de sodium pH 4,5 (vert) ; bicine pH 9 (violet) et Hepes pH 7 (rouge), dans un tampon contenant également 150 mM de NaCl et 2% de glycérol. (B) Analyses SDS-PAGE 10% colorés au bleu de Coomassie de PII1-H2H3 après purification par chromatographie d'exclusion stérique. La flèche correspond à la région de migration attendue pour PII1-H2H3.

Chaque profil d'élution présente de nombreux pics imbriqués quel que soit le tampon utilisé. On peut regrouper ces pics en fonction du volume d'élution : à 7, 12 et 22 ml qui sont encadrés en rouge, vert ou bleu respectivement sur la **Figure 31A**. L'intensité des pics à 7 et 12 ml est inférieure à 50 mAU, ce qui peut se traduire par une faible quantité de protéines. PII1 possède peu de résidus aromatiques dont un seul tryptophane, ce qui peut expliquer la faible valeur d'absorbance. Contrairement à ce à quoi on pourrait s'attendre, les protéines éluées à 7 ml sont visibles sur SDS-PAGE et correspondent à PII1-H2H3 et cela dans tous les tampons, sauf dans l'acétate de sodium, où PII1-H2H3 n'est pas stable (**Figure 31B**). Pour un temps de rétention plus élevé (12 ml), on constate que ce sont surtout les protéines d'*E. coli* qui sont éluées. Au volume de rétention de 22 ml, aucune protéine n'est visible sur SDS-PAGE.

Finalement, le pH ne semble pas avoir d'effet sur la séparation des pics en revanche on peut observer que dans l'HEPES, on récupère un peu plus de protéines en comparaison aux autres tampons et la bande correspondant à PII1-H2H3 est plus nette. Pour ces raisons nous avons sélectionné le tampon HEPES à pH 7 pour continuer les études.

En résumé de ce chapitre, nous avons testé plusieurs méthodes d'expression et de purification pour PII1 ou PII1-H2H3. L'ensemble des résultats obtenus montre que PII1 entière est très peu soluble et stable ce qui ne permet pas de l'utiliser par les approches structurales classiques (cristallographie aux RX ou Small Angle X-rays Scattering SAXS). La présence de régions CC est probablement responsable d'interactions non spécifiques des molécules entre elles, ce qui rend l'étape de purification difficile. Les constructions tronquées que j'ai réalisées permettent néanmoins d'obtenir des protéines solubles, que nous avons pu purifier par chromatographie d'affinité. Si l'étape de chromatographie SEC ne permet ni l'étude de la polydispersité de la protéine ni d'évaluer son état oligomérique, elle a néanmoins permis d'éliminer quelques contaminants et nous avons gardé cette étape dans le protocole de purification pour l'étude SR-CD. Les expériences de doubles hybrides avec SS4 comme protéine proie (Vandromme et al., 2019) avaient montré une interaction avec PII1. Ces observations nous ont conduit à penser que la stabilité de PII1 nécessite peut-être la présence d'un ou plusieurs de ses partenaires pour acquérir une stabilité en solution. Nous avons donc entrepris, dans un premier temps, la production de son partenaire SS4 qui joue un rôle majeur dans l'initiation des grains d'amidon.

Chapitre 2. Analyse de l'expression de SS4

Dans ce chapitre nous avons entrepris la production de SS4 pour, à terme, coexprimer PII1 (et constructions tronquées) et SS4. En effet, en raison de l'instabilité de PII1 nous avons envisagé de stabiliser PII1 par l'interaction avec son partenaire SS4. Les résultats de la co-expression seront décrits dans le chapitre 3. Dans ce chapitre 2, nous avons utilisé trois constructions tronquées de SS4 plus faciles à produire que la protéine entière, moins stable. Ces clonages ont été réalisés avec l'aide du Dr David Dauvillée (USGF, Lille). Ce fut le fruit d'un long travail qui a nécessité beaucoup de temps étant donné que nous avons utilisé les enzymes de restriction plutôt que le système Gateway pour cloner. J'ai également eu de la difficulté à obtenir des séquences nucléotidiques sans mutation après la ligation. Après avoir obtenu les vecteurs plasmidiques d'intérêt nous avons par la suite entamé l'expression et la purification des échantillons protéiques des constructions de SS4.

2.1 Clonage des constructions tronquées SS4

Dans notre équipe de recherche, des tests d'expression ont été menés sur la forme entière de SS4 (110 kDa). Les résultats montraient que SS4 subissait une protéolyse, visible dans la fraction soluble (**Figure 32**). En effet, on remarque la présence de deux bandes autour de 55 kDa qui ont pu être identifiées par spectrométrie de masse et correspondant à deux fragments stables de SS4 (C-terminal et N-terminal). Ces fragments semblent issus d'une coupure environ au milieu de la séquence (1040 résidus).



Figure 32. Analyse SDS-PAGE de la purification par chromatographie d'affinité de SS4 pDest17 produite dans BL21(DE3). Résultats issus de la thèse de Maude Facon (USGF, 2014).

À partir de ce constat, j'ai stratégiquement effectué des constructions tronquées de SS4 plutôt que de travailler sur la protéine entière pour améliorer les chances de solubilité (comme ce fut le cas avec PII1). Pour sélectionner les constructions tronquées, je me suis basée sur les travaux de Raynaud et al.,(2016) qui proposent que SS4 soit formée d'une région N-terminale composée de quatre régions CC (CC1, CC2, CC3 et CC4) et une région de dimérisation (RD) (cf. Introduction, Chapitre 3.1.4 Rôles distincts des domaines de SS4) et d'une région C-terminale comprenant le domaine catalytique. Dans cette étude, SS4 a été exprimée dans *E. coli* sous différentes versions tronquées solubles avec la région CC3 jusqu'au C-terminal ou la région RD jusqu'au C-terminal (Raynaud et al., 2016) (cf BioRxiv Bossu et al., 2024).

Le PM calculé à partir de ProtParam de la région N-terminale et C-terminale fait respectivement 61 kDa et 56 kDa et coïncide avec la taille des fragments observées sur gel SDS-PAGE, d'environ 55 kDa (**Figure 32**). Cela signifie que ces deux régions sont stables et solubles séparément en solution. De plus, la structure du domaine catalytique ayant été résolue par cristallographie, nous sommes partis du constat que la partie catalytique serait stable et soluble (Nielsen et al., 2018).

A partir de ces observations, nous avons donc utilisé l'extrémité C-terminale, comme point de départ des constructions, puis rajouté progressivement des régions prédites en CC du N-terminal (CC3 ou CC4), dans l'optique de pouvoir former des complexes avec PII1. Enfin nous avons conservé la région RD qui est nécessaire à la dimérisation de SS4 et car c'est sous la forme dimérique qu'elle interagit avec PII1 et que son activité catalytique est fonctionnelle.

En résumé, nous avons réalisé trois constructions tronquées de SS4 clonées dans un vecteur pETDuet, qui permet l'ajout d'une étiquette 6-His en N-terminal (**Figure 33**). La première construction contient le domaine catalytique et le domaine de dimérisation (SS4Δ466, 63 kDa). La deuxième construction contient en supplément le CC4 (SS4Δ403, 69 kDa) et la dernière construction contient CC3, CC4, RD et le domaine catalytique (SS4Δ349 75 kDa)(**Figure 33**).



Figure 33. Description schématique des constructions tronquées de SS4. Figure du haut : Structure de SS4 décrite par Raynaud et al 2007, les régions coiled-coils sont indiquées en orange, la région de dimérisation (CR) en violet et la région catalytique en bleu. Figure du bas : Structure des constructions tronquées de SS4 réalisées. TP=peptide de transite, GT= glycosyltransférase.

2.2 Tests d'expression des constructions tronquées de SS4

Les tests d'expression ont été réalisés avec le même protocole que celui utilisé pour les constructions de PII1 décrit dans la section « Matériel et méthodes ». Les conditions testées sont listées dans les **Tableaux 7, 8 et 9**. Les résultats sont montrés **Figure 34**. Le PM attendu des constructions tronquées est de 75 kDa pour SS4Δ349, 69 kDa pour SS4Δ403 et 63 kDa pour SS4Δ466.



Figure 34. Tests d'expression des constructions tronquées SS4 pETDuet. Le mode d'induction est indiqué sous les gels. NI= fraction non induite, I = fraction induite, S = fraction soluble. Les flèches correspondent à la zone de migration attendue pour les constructions tronquées de SS4.

Les résultats montrent que les trois constructions sont exprimées (Piste I) et solubles (Piste S), bien que retrouvées en petite quantité. On observe que pour chaque construction, l'induction à 20°C sur la nuit de même que la lyse dans le tampon HEPES (à l'exception de Δ 466, plus soluble dans le tampon MES), permet d'obtenir le plus de protéines solubles.

Afin de visualiser plus facilement les protéines sur SDS-PAGE, nous avons concentré très fortement les protéines à partir d'une culture de 100 ml dans 50 µl de résine et éluées avec 50 µl de tampon contenant 500 mM d'imidazole (**Figure 35**).



Figure 35. Purification et caractérisation de SS4Δ349, SS4Δ403 et SS4Δ466. (A) SDS-PAGE 10% des constructions tronquées de SS4 pETDUET produites dans BL21(DE3), purifiées sur résine de Nickel.(B) Western-blot anti-His des constructions tronquées SS4 à partir des échantillons purifiés par affinité sur résine de nickel.

Les résultats montrent que les protéines sont solubles et purifiées après une étape qui permet de se débarrasser de nombreux contaminants et de les distinguer visuellement des protéines de *E. coli*. Pour confirmer la présence de SS4, nous avons réalisé un Westernblot anti-6-His sur ces mêmes échantillons (**Figure 35B**). L'analyse du gel Western-Blot confirme la présence des protéines sous forme soluble. On observe que la bande de SS4Δ466 est moins intense que celle des autres constructions et cela s'explique par le fait que l'étiquette 6-His est probablement moins accessible aux anticorps anti-6-His.

2.3 Purification par chromatographie d'exclusion stérique

À la suite de la purification par chromatographie d'affinité sur résine par batch à partir d'une culture de 100 ml dans le milieu LB, nous avons souhaité ajouter une deuxième étape de purification afin de se débarrasser des quelques contaminants restants. Afin de récupérer un maximum de protéines pures pour l'analyse de SS4, nous avons réalisé des cultures de 4x 500 ml. Étant donné que la purification d'affinité a permis d'obtenir des échantillons concentrés et relativement purs nous avons gardé le même protocole que précédemment en ajoutant proportionnellement la quantité de résine au volume de culture, soit 1 ml de résine pour 2L de culture. L'échantillon purifié a été déposé sur SDS-PAGE puis injecté directement dans la colonne de purification par exclusion stérique Superdex 200 10/300 GL. Les résultats sont présentés dans la **Figure 36**.



*Figure 36. Purification des constructions tronquées SS4 pETDUET par chromatographie d'exclusion stérique sur colonne Superdex 200 10/300 GL. Figure de gauche : profils d'élution superposés de SS4*Δ349 (bleu), SS4Δ403 (orange) et SS4Δ466 (vert). Les flèches colorées indiquent les pics dans lesquels sont retrouvées les constructions tronquées de SS4. Figures de droite : SDS-PAGE 10 % colorés au bleu de Coomassie des protéines purifiées. Les pistes déposées correspondent aux pics des profils d'élutions. La piste affinité correspond à l'échantillon issus de l'élution de la première étape et injecté dans la colonne.

Les profils d'élution montrent plusieurs pics relativement superposés. On peut compter 6 pics majoritaires à des volumes de rétention allant de 7 à 20 ml et d'intensité variable. Par ailleurs, on remarque un décalage du pic 3 selon les constructions protéiques. Sur SDS-PAGE, le pic 3 correspond aux protéines SS4Δ349 (Pistes 3 et 4), SS4Δ403 (Piste 3) et SS4Δ466 (Piste 4). Les protéines sont éluées à 12 ml.

Cependant, on observe une faible quantité de protéines présentes après la deuxième étape de purification, et les échantillons contiennent encore des contaminants dont il est difficile de se débarrasser. Les fractions injectées après la purification réalisée par chromatographie d'affinité contiennent moins de protéines d'intérêt et plus de contaminants par rapport à la purification dans un petit volume de culture (100 ml) (**Figure 35A**). La DO d'induction des protéines exprimées dans 100 ml était supérieure à celle des protéines exprimées dans un grand volume et pourrait expliquer cet écart. De plus, le lavage de la résine n'avait pas été fait avec 25 mM d'imidazole et a probablement eu un impact sur
la pureté de l'échantillon. Dans l'optique de récupérer un maximum de protéines pures, j'ai purifié les échantillons sur résine en prenant soin de laver correctement la résine de Nickel (résultats non présentés). Malgré l'amélioration de la pureté après la première étape de purification, cela a entraîné une perte de la quantité de protéines lors de la deuxième étape de purification, de telle sorte que la quantité de protéines récupérées n'était pas satisfaisante.

En résumé, la production des protéines seules montre que les constructions tronquées de SS4 sont solubles et permettent de commencer les études de complexes avec PII1 toutefois la purification est difficile et comme PII1-H2H3, la protéine co-purifie avec des contaminants. En conclusion, de la même façon que pour PII1-H2H3, SS4 est difficile à purifier seule car elle n'est pas très stable sans ses partenaires de l'initiation. Il est probable que la suppression des CC1 et CC2 affecte également cette stabilité.

Chapitre 3. Formation du complexe PII1/SS4

Dans ce chapitre 3, le dernier qui concerne l'expression des protéines, nous allons nous attarder sur l'expression de PII1 et SS4 en mélange à partir des échantillons protéiques exprimés séparément ou co-exprimés.

3.1 Co-purification des constructions tronquées de SS4 et PII1

3.1.1 Co-purification par affinité

Les résultats obtenus dans le chapitre 1 nous ont montré que PII1 était très peu soluble en solution et que les constructions PII1∆71 et PII1∆101 étaient solubles mais en trop faible quantité pour des études structurales. En revanche, la construction PII1-H2H3 sur laquelle nous avions réalisé beaucoup de tests de purification est soluble, bien que difficile à purifier, avec un fort degré de pureté. Ayant émis l'hypothèse que PII1 pourrait être plus stable et soluble en solution en présence de SS4 (et inversement) j'ai réalisé des essais de co-purification de la protéine PII1 entière ou tronquée avec les constructions tronquées SS4 pour vérifier si la co-purification permettrait d'améliorer la solubilité et la stabilité des protéines impliquées dans un complexe.

Pour cela, les protéines ont été exprimées séparément dans la souche BL21(DE3) à partir d'une culture de 100 ml. Ensuite les culots bactériens ont été resuspendu dans le même tampon de lyse et mélangés avant lyse. L'extrait soluble du mélange est ensuite soumis à une étape d'affinité par batch en rajoutant 50 µl de résine de Nickel et les protéines sont éluées dans 50 µl de tampon contenant 500 mM d'imidazole (**Figure 37**). Enfin les protéines PII1 et PII1-H2H3 en mélange avec les protéines SS4 ont été identifiées par Western-Blot anti-6-His afin de confirmer leur présence et voir si potentiellement le mélange avec SS4 pouvait aider à la solubilisation de PII1 entière. L'ensemble des co-purifications par affinité entre les constructions de PII1 et de SS4 sont résumés dans les **Tableaux 10 et 11**.



Figure 37. Expression des constructions tronquées PII1 et SS4 co-purifiées. (A) SDS-PAGE 10% colorés au bleu de Coomassie de la co-purification sur résine de Nickel entre PII1 (entière, PII1 Δ 71, PII1 Δ 101, PII1-H1H2H3, PII1-H2H3) et SS4 Δ 349 (vert), SS4 Δ 403(bleu) ou SS4 Δ 466 (jaune)). **B**) Western Blot anti-6-His des protéines issues de (A). Les flèches noires = PII1 (entière ou constructions tronquées). Les flèches vertes, bleues ou jaunes = SS4 Δ 349, SS4 Δ 403 ou SS4 Δ 466 respectivement. I=échantillon protéique induit PII1. E = échantillon protéique PII1 purifié seul.

Les résultats montrent que les trois constructions tronquées de SS4 sont exprimées en quantités équivalentes. La présence des constructions tronquées de SS4 est vérifiée par un Western blot anti-6-His, qui les identifie correctement contrairement à PII1 entière. Concernant les protéines PII1, on n'observe aucune bande (ou seulement très faiblement visible) pour PII1 (90 kDa), PII1 Δ 71 (82 kDa) et PII1 Δ 101 (78 kDa) ce qui indique que la copurification des protéines avec SS4 n'a pas permis d'améliorer la solubilité par rapport à la purification des protéines seules. Enfin, le mélange de PII1-H2H3 ou PII1-H1H2H3 et les protéines SS4 tronquées permet de récupérer une quantité relativement homogène de protéines solubles. La présence des protéines PII1-H2H3 et SS4 a été confirmée par le Western Blot anti-6-His (**Figure 37B**).

À la suite de ces résultats, nous avons décidé de travailler essentiellement sur la construction de PII1-H2H3 en complexe avec la construction SS4Δ349 ou SS4Δ403, d'une part pour se concentrer sur la purification des échantillons les plus solubles et d'autre part pour récupérer une quantité suffisante de protéines afin d'effectuer la deuxième étape de purification, qui permettra de vérifier l'interaction des protéines.

3.1.2 Co-purification par chromatographie d'exclusion stérique

La première étape par chromatographie d'affinité ne permettant pas de vérifier la formation d'un complexe protéique, étant donné que les deux protéines PII1-H2H3 et SS4Δ349/SS4Δ403 contiennent une étiquette 6-His, nous avons donc ajouté une deuxième étape de purification par chromatographie d'exclusion stérique dans une colonne Superdex 200 10/300 GL à partir d'une culture de 500 ml. Cette étape devrait permettre de vérifier l'interaction entre les deux protéines et en parallèle de se débarrasser des contaminants (**Figure 38**). Toutes les purifications effectuées lors des co-purifications sont répertoriées dans les **Tableaux 10 et 11**.



Figure 38. Co-purification de PII1-H2H3 et SS4Δ349 (A,B,C) ou PII1-H2H3 et SS4Δ403 (D,E) par chromatographie d'exclusion stérique sur colonne Superdex 200 10/300 GL. (A) Profil d'élution du mélange PII1-H2H3 et SS4Δ349. (B) SDS-PAGE 10% coloré au bleu de Coomassie. La piste NI2+ correspond à l'échantillon injecté. (C) Western-blot anti-6-His des pistes 3 et 9 de (B). (D) Profil d'élution du mélange PII1-H2H3 et SS4Δ403. (E) SDS-PAGE 10% coloré au bleu de Coomassie. La piste NI2+ correspond à l'échantillon injecté. La flèche correspond à la zone de migration attendue pour PII1-H2H3 et SS4.

Le profil d'élution du mélange PII1-H2H3 et SS4 Δ 349 présente six pics peu séparés. Les protéines sont éluées à un volume de rétention allant de 7 à 40 ml soit près de deux fois le volume de colonne (**Figure 38A**). Le SDS-PAGE permet d'identifier PII1-H2H3 seule dans les pistes 1 à 3 et en mélange avec SS4 Δ 349 dans toutes les autres pistes. En se référant au chromatogramme on observe que PII1 est éluée surtout dans le pic 1 à 10 ml environ. SS4 Δ 349 est visible dans les pics 2 et 3 et est éluée entre 15 et 20 ml. Etant donné que les pics ne sont pas bien séparés, il est difficile d'identifier un volume de rétention fiable et donc un poids moléculaire. En revanche on observe la présence des deux protéines dans les pics 2 et 3 qui met en évidence l'interaction entre PII1-H2H3 et SS4Δ349. Le Western Blot anti- 6- His confirme la présence de PII1-H2H3 et SS4Δ349 (faiblement visible) et donc leur interaction.

De la même façon qu'avec SS4 Δ 349, nous avons co-purifié le mélange de PII1-H2H3 et SS4 Δ 403. Les analyses SDS-PAGE montrent que PII1-H2H3 seule est éluée majoritairement dans la piste 2 correspondant au pic 2 à 8,5 ml, bien qu'elle soit présente également en quantité plus faible dans toutes les autres pistes (**Figure 38C**). La protéine SS4 Δ 403 seule est éluée majoritairement dans la piste 6 à 13 ml. Cette purification confirme également l'interaction *in vitro* entre PII1-H2H3 et SS4 Δ 403.

Pour la suite du travail nous avons choisi de travailler avec SS4∆349 qui est la version qui se rapproche le plus de SS4 entière.

En résumé, nous avons pu démontrer la capacité de PII1-H2H3 à interagir avec SS4 Δ 349 et SS4 Δ 403 en solution. Ces résultats encourageants montrent également que malgré la suppression des CC1 et CC2 (et CC3 dans le cas de SS4 Δ 403), SS4 est capable d'interagir avec son partenaire PII1-H2H3. Cela permet également de montrer que les hélices H2 et H3 de PII1 suffisent pour l'interaction et donc précise les régions d'interaction. Toutefois, la production du mélange de PII1-H2H3 et SS4 Δ 349 ne permet pas d'obtenir un échantillon protéique homogène. Afin d'obtenir un échantillon plus stable du complexe PII1/SS4, nous avons opté pour une autre stratégie en construisant un double clonage permettant la co-expression de PII1 et SS4 Δ 349 au cas où les protéines nécessiteraient d'être exprimées ensemble pour pouvoir être stables en solution.

3.2 Co-expression de PII1 et SS4

3.2.1 Purification du complexe par chromatographie d'affinité

Nous avons élaboré une autre stratégie pour obtenir un complexe protéique stable. Cette fois-ci nous avons cloné la protéine PII1 entière avec la construction SS4∆349 dans le vecteur pETDuet dans l'espoir que PII1 puisse se stabiliser par la présence de SS4Δ349 (et inversement) au cours de leur repliement. De plus, nous comptons sur ce complexe pour augmenter le rendement protéique et améliorer l'interaction entre les protéines

Lors de ma thèse, j'ai également travaillé sur l'obtention de double clonage entre PII1 et SS4∆403 ou SS4∆466 mais seule la construction exprimant le complexe PII1- SS4∆349 a pu être obtenue. Ce clonage a été le fruit de plusieurs mois de travail réalisés avec l'aide du Dr David Dauvillée à l'UGSF.

Dans ce vecteur d'expression, seule PII1 possède une étiquette 6-His. L'idée étant de pouvoir vérifier facilement l'interaction entre les deux protéines dès la première étape de purification par chromatographie d'affinité. Les protéines ont été produites dans 500 ml de culture, dans la souche d'expression BL21 (DE3) et induites par ajout de 0.5 mM d'IPTG pendant toute une nuit à 20°C. Les protéines solubles d'intérêt sont récupérées après la lyse des bactéries et centrifugation. L'échantillon soluble a été incubé avec 500 µl de résine de nickel et les protéines ont été éluées en faisant un gradient d'imidazole de 50 à 200 mM (Figure 39). L'ensemble des tests d'expression effectués sur le complexe est résumé dans le Tableau 12.



Figure 39. Purification du complexe PII1-SS4Δ349 par affinité sur résine de Nickel. SDS-PAGE 10% coloré au bleu de Coomassie du complexe PII1-SS4Δ349 pETDuet exprimé dans BL21(DE3). Les protéines ont été éluées par un gradient croissant de 50 à 200 mM d'imidazole. NI = fraction non induites et I =fraction induite, S = fraction soluble et FT = fraction non retenue. Les flèches correspondent à la zone de migration attendue pour le complexe.

Les résultats montrent que PII1 (90 kDa) est exprimée en quantité visible dans la fraction induite. En revanche, comme pour les protéines exprimées séparément, SS4

(70 kDa) est peu exprimée dans la fraction induite mais on repère plus facilement sa présence dans la fraction soluble. Après la purification à 50 et 100 mM d'imidazole, on observe deux bandes au PM attendu pour PII1 et SS4Δ349 pouvant correspondre à un complexe entre les deux protéines On observe également une bande à 50 kDa qui doit être probablement une protéine issue d'*E. coli*.

Pour vérifier la présence du complexe PII1/SS4Δ349 une deuxième étape de purification par chromatographie d'exclusion stérique a été réalisée.

3.2.2 Purification du complexe par chromatographie d'exclusion stérique

Les fractions d'élutions issues de la première étape de chromatographie d'affinité ont été concentrées puis injectées sur colonne de gel filtration Superdex 200 10/300 GL (**Figure 40**).



Figure 40. Purification du complexe PII1-SS4∆349 par chromatographie d'exclusion stérique Superdex 200 10/300 GL. (A) Profil d'élution du complexe. (B) SDS-PAGE 10% coloré au bleu de Coomassie des protéines purifiés. La piste NI2+ correspond à l'échantillon injecté. Les pistes 1 à 8 correspondent aux pics issus du profil d'élution (A). Les flèches correspondent à la zone de migration attendue pour le complexe.

Comme dans le cas des protéines exprimées seules, on observe un profil d'élution comportant des pics peu résolus entre 7 et 20 ml. Sur SDS-PAGE, des bandes au PM attendu pour PII1 sont visibles dans les pistes 4, 5 et 6 et est éluée à 12 ml, et des bandes correspondant au PM de SS4 Δ 349 éluées à 11 ml, dans les pistes 3 et 4. Les bandes correspondant aux deux protéines sont présentes dans les pistes 4, 5 et 6. Pour confirmer la présence des protéines nous avons soumis les bandes correspondantes aux protéines en spectrométrie de masse LC-MS/MS, (Fabrice Bray, MSAP, Villeneuve d'Ascq). Malheureusement, malgré plusieurs essais la présence de PII1 n'a pas pu être identifiée et seulement quelques peptides de SS4 Δ 349 ont été identifiés avec ceux correspondant à des protéines bactériennes.

En conclusion de ce chapitre, nous avons pu démontrer que PII1 et SS4 interagissent, ce qui est un résultat très satisfaisant bien que le complexe n'ait pas pu être purifié. Du fait de la difficulté rencontrée pour ce double clonage, du peu de fiabilité du résultat obtenu et du peu de temps restant pour ma thèse, et suivant les conseils prodigués lors de mon CSI, j'ai décidé de ne pas continuer les expériences de co-expression et de me concentrer sur une étude structurale par modélisation moléculaire et sur l'expression des protéines individuelles pour des tests fonctionnels et une étude en dichroïsme circulaire sur PII1-H2H3. En effet, ces approches, débutées en parallèle des essais sur la production de protéines recombinantes, donnaient des résultats plus encourageants pour l'étude de ces protéines.

Tableau 4. Conditions d'expression et de purification utilisées pour l'expression PII1 entière. AI = auto-induction, SEC= chromatographie d'exclusion stérique, IEX = chromatographie échange d'anions.

Protéine	Souche	Plasmide	Condition induction	additifs	Lyse	Batch/Colonne	SEC/IEX	Commentaire
PII1	BL21 (DE3)	pML375	0,5 mM IPTG 20°C ON		bper	Batch Ni-NTA		
PII1	BL21 (DE3)	pML375	0,1 mM IPTG 20°C ON		bper	Batch Ni-NTA		
PII1	Rosetta	pML375	0,1 mM IPTG 20°C ON		bper	Batch Ni-NTA		
PII1	Rosetta	PML375	0,5 mM IPTG 20°C ON		bper	Batch Ni-NTA		
PII1	BL21 AI	pET300	0,1mM IPTG 20°C ON		bper	Batch Ni-NTA		
PII1	BL21 AI	pET300	AI 20°C ON		bper	Batch Ni-NTA		
PII1	BL21 AI	PML375	0,1mM IPTG 20°C ON		bper	Batch Ni-NTA		
PII1	BL21 AI	PML375	0,5 mM IPTG 20°C ON		bper	Batch Ni-NTA		
PII1	BL21 (DE3)	pET300	AI 20°C / 15°C ON		bper	Batch Ni-NTA		
PII1	BL21 (DE3)	pET300	AI 20/30/37°C ON	glucose/lactose/glycérol	bper	Batch Ni-NTA		Criblage milieu auto-inductible
PII1	BL21 (DE3)	pET300	0,1 mM IPTG 20°C ON		bper	Batch Ni-NTA		
PII1	C41	pET300	AI 20°C ON		bper	Batch Ni-NTA		
PII1	C43	pET300	AI 20°C ON		bper	Batch Ni-NTA		
PII1	Rosetta G	pET300	AI 20°C ON		bper	Batch Ni-NTA		
PII1	Rosetta G	pML333	AI 20°C ON		bper	Batch Ni-NTA		
PII1	BL21 (DE3)	pML333	AI 20°C ON		bper	Batch Ni-NTA		
PII1	BL21 (DE3)	pML310	AI 20°C ON		bper	Batch Ni-NTA		
PII1	BL21 (DE3)	pET300	AI 20°C ON		Emulsiflex	Batch Ni-NTA		
PII1	BL21 (DE3)	pET300	AI 20°C ON	0,1 mM DTT	Emulsiflex	Batch Ni-NTA		
PII1	BL21 (DE3)	pET300	AI 20°C ON	Triton 0,1X	Emulsiflex	Batch Ni-NTA		
PII1	BL21 (DE3)	pET300	AI 20°C ON	15mM EDTA	Emulsiflex	Batch Ni-NTA		
PII1	BL21 (DE3)	pET300	AI 20°C ON	1M NaCl	Emulsiflex	Batch Ni-NTA		
PII1	BL21 (DE3)	pET300	AI 20°C ON		Emulsiflex	Batch Ni-NTA	SEC 25 ml	
PII1	BL21 (DE3)	pET300	AI 20°C ON		Emulsiflex		IEX 1 ml	
PII1	BL21 (DE3)	pET300	AI 20°C ON		Emulsiflex	Hi trap 1 ml + dessalage	IEX 1 ml	
PII1	BL21 (DE3)	pET300	AI 20°C ON	2M urée	Emulsiflex			Solubilisation corps d'inclusion

Tableau 5. Conditions d'expression et de purification utilisées pour PII1-H2H3. AI = auto-induction, SEC= chromatographie d'exclusion stérique, IEX = chromatographie échange d'anions.

Protéine	Souche	Plasmide	Conduction induction	additifs	Lyse	Batch/Colonne Ni	SEC/IEX
PII1-H2H3	BL21 (DE3)	pET300	0,1 mM IPTG 20°C ON		sonication	Batch Ni-NTA	
PII1-H2H3	BL21 (DE3)	pET300	0,5 mM IPTG 20°C ON		sonication	Batch Ni-NTA	
PII1-H2H3	BL21 (DE3)	pET300	1 mM IPTG 20°C ON		sonication	Batch Ni-NTA	
PII1-H2H3	BL21 (DE3)	pET300	0,1 mM IPTG 37°C 3h		sonication	Batch Ni-NTA	
PII1-H2H3	BL21 (DE3)	pET300	0,1 mM IPTG 16°C ON		sonication	Batch Ni-NTA	
PII1-H2H3	BL21 (DE3)	pET300	AI 20°C ON		sonication	Batch Ni-NTA	
PII1-H2H3	BL21 (DE3)	pET300	AI 20°C ON		sonication	His Trap Excel 5 ml	SEC 25ml+ IEX 1ml
PII1-H2H3	BL21 (DE3)	pET300	AI 20°C ON	4 mM DTT 5 mM EDTA	emulsiflex	col 1 ml + dess	IEX 1ml
		FT000				Batch Ni-	
PII1-H2H3	BL21 (DE3)	pE1300	AI 20°C ON		emulsiflex	NIA	SEC 25 ml
	BL01 (DE2)	- ET200	4L209C ON		a mulaiflav	Batch Ni-	
РШ-п2пэ	BL21 (DE3)	pE1300	AI 20°C ON	0,1 MM D110,1 MM EDIA	emuisinex	NIA+ dess	
PII1-H2H3	BI 21 (DE3)	pET300	AL20°C ON		emulsiflex	Batch Ni- NTA + dess	SEC 25 ml
111112110	BE21 (BE0)	percoo	1120 0 011		ematomex	Patah Ni	02020111
PII1-H2H3	BL21 (DE3)	pET300	AI 20°C ON		emulsiflex	NTA+ dess	SEC 25 ml
	,					Batch Ni-	
PII1-H2H3	BL21 (DE3)	pET300	AI 20°C ON		emulsiflex	NTA+ dess	SEC 25 ml
						Batch Ni-	
PII1-H2H3	BL21 (DE3)	pET300	AI 20°C ON		emulsiflex	NTA+ dess	SEC 25 ml
						batchNi-	
PII1-H2H3	BL21 (DE3)	pET300	AI 20°C ON		emulsiflex	NTA+ dess	SEC 120 ml
PII1-H2H3	BL21 (DE3)	pET300	AI 20°C ON		emulsiflex		IEX 5 ml + SEC 120 ml
PII1-H2H3	BL21 (DE3)	pET300	AI 20°C ON	0,1 mM DTT	Sonication		IEX 5 ml + SEC 120 ml
						Batch Ni-	
PII1-H2H3	BL21 (DE3)	pET300	AI 20°C ON	10% acétonitrile	emulsiflex	NTA+ dess	SEC 25 ml
	BI 21 (DE2)	pET300		0.1 mM DTT 0.1 mM EDTA	omulaifley	Batch Ni- NTA	SEC 25 ml
PII1_H2H3	BL 21 (DE3)	pE1300	0.5 mM IPTG	o, miniori o, minicula	emulsiflev	col 1 ml	SEC 25 ml

Tableau 6. Conditions d'expression et de purification utilisées pour PII1 ∆**71, PII1** ∆**101 et PII1-H1H2H3.** AI = auto-induction, SEC= chromatographie d'exclusion stérique, IEX = chromatographie échange d'anions.

Protéine	Souche	Plasmide	Conduction induction	Lyse	Batch/Colonne Ni
PII1∆71	BL21 (DE3)	pET300	0,1 mM IPTG 20°C ON	sonication	Batch Ni-NTA
PII1∆71	BL21 (DE3)	pET300	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA
PII1∆71	BL21 (DE3)	pET300	AI 20°C ON	sonication	Batch Ni-NTA
PII1∆71	BL21 (DE3)	pET300	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA

Protéine	Souche	Plasmide	Conduction induction	Lyse	Batch/Colonne Ni	SEC/IEX
PII1-H1-H2-H3	BL21 (DE3)	pET300	0,1 mM IPTG 20°C ON	sonication	Batch Ni-NTA	
PII1-H1-H2-H3	BL21 (DE3)	pET300	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA	
PII1-H1-H2-H3	BL21 (DE3)	pET300	AI 20°C ON	sonication	Batch Ni-NTA	
PII1-H1-H2-H3	BL21 (DE3)	pET300	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA	
PII1-H1-H2-H3	BL21 (DE3)	pET301	0,5 mM IPTG 20°C ON	sonication	His Trap Excel 5 ml	
PII1-H1-H2-H3	BL21 (DE3)	pET300	0,1 mM IPTG 20°C ON	emulsiflex	His Trap Excel 5 ml	SEC 25 ml
PII1-H1-H2-H3	BL21 (DE3)	pET300	AI IPTG 20°C ON	sonication	Batch Ni-NTA	SEC 25 ml

Protéine	Souche	Plasmide	Conduction induction	Lyse	Batch/Colonne Ni
PII1∆101	BL21 (DE3)	pET300	0,1 mM IPTG 20°C ON	sonication	Batch Ni-NTA
PII1∆101	BL21 (DE3)	pET300	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA
PII1∆101	BL21 (DE3)	pET300	AI 20°C ON	sonication	Batch Ni-NTA
PII1∆101	BL21 (DE3)	pET300	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA

Tableau 7 Conditions d'expression et de purification utilisées pour la construction $SS4 \land 34$	9 SEC= chromatographie d'exclusion stérique Al= auto-induction

					Batch/Colonne	
Protéine	Souche	Plasmide	Condition induction	Lyse	Ni	SEC/IEX
SS4∆349	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA	
SS4∆349	BL21 (DE3)	pETDuet	AI 20°C ON	sonication	Batch Ni-NTA	
SS4∆349	BL21 (DE3)	pETDuet	1mM IPTG 1h30/ 2h15/3h 37°C			
SS4∆349	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	bper		
SS4∆349	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆349	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆349	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆349	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆349	Rosetta	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆349	Rosetta	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆349	Rosetta	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆349	Rosetta	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆349	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	Emulsiflex	Batch Ni-NTA	SEC 25ml
SS4∆349	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	Emulsiflex	+ batch Ni-NTA	SEC 25ml

Tableau 8. Conditions d'expression et de purification utilisées pour la construction SS4 403. SEC= chromatographie d'exclusion stérique, AI= auto-induction

Protéine	Souche	Plasmide	Condition induction	Lyse	Batch/Colonne Ni	SEC/IEX
SS4∆403	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA	
SS4∆403	BL21 (DE3)	pETDuet	AI 20°C ON	sonication	Batch NI-NTA	
			1mM IPTG 1h30/ 2h15/3h			
SS4∆403	BL21 (DE3)	pETDuet	37°C			
SS4∆403	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	bper		
SS4∆403	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆403	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆403	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆403	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆403	Rosetta	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆403	Rosetta	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆403	Rosetta	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆403	Rosetta	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆403	BL21 (DE3)	pETDuet	1 mM IPTG 20°C ON	Emulsiflex	His Trap Excel 1 ml	SEC 25 ml
SS4∆403	BL21 (DE3)	pETDuet	1 mM IPTG 20°C ON	Emulsiflex	Batch Ni-NTA	SEC 25 ml
SS4∆403	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA	SEC 25ml

Tableau 9. Conditions d'expression et de purification utilisées pour la construction SS4 (2466. SEC = chromatographie d'exclusion stérique, AI = auto-induction.

Protéine	Souche	Plasmide	Condition induction	Lyse	Batch/Colonne N	i SEC/IEX
SS4∆466	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA	
SS4∆466	BL21 (DE3)	pETDuet	AI 20°C ON	sonication	Batch Ni-NTA	
SS4∆466	BL21 (DE3)	pETDuet	1 mM IPTG 1h30/ 2h15/3h 37°C	;		
SS4∆466	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	bper		
SS4∆466	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆466	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆466	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆466	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆466	Rosetta	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆466	Rosetta	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆466	Rosetta	pETDuet	0,5 mM IPTG 20°C ON	sonication		
SS4∆466	Rosetta	pETDuet	0,5 mM IPTG 20°C ON	sonication		
					His Trap Excel 1	SEC 25
SS4∆466	BL21 (DE3)	pETDuet	1 mM IPTG 20°C ON	Emulsiflex	ml	ml
					His Trap Excel 5	SEC 25
SS4∆466	BL21 (DE3)	pETDuet	1 mM IPTG 20°C ON	Emulsiflex	ml	ml
						SEC
SS4∆466	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA	25ml

Tableau 10. Conditions d'expression et de purification utilisées pour les co-purifications entre PII1-H2H3 et les constructions tronquées de PII1 et SS4. SEC= chromatographie d'exclusion stérique. IEX= chromatographie par échange d'anions. Dess = dessalage.

Protéine	Souche	Plasmide	Condition induction	Lyse	Batch/Colonne Ni	SEC/IEX
PII1-H2H3+SS4∆349	BL21 (DE3)	pET300+ pETDuet	0,5 mM IPTG 20°C ON	emulsiflex		
PII1-H2H3+SS4∆349	BL21 (DE3)	pET300+ pETDuet	0,5 mM IPTG 20°C ON	sonication	batch Ni-NTA+ dess	IEX 1 ml
PII1-H2H3+SS4∆349	BL21 (DE3)	pET300+ pETDuet	0,5 mM IPTG 20°C ON	emulsiflex	His Trap Excel 5ml	SEC 25 ml
PII1-H2H3+SS4∆349	BL21 (DE3)	pET300+ pETDuet	0,5 mM IPTG 20°C ON	Sonication	Batch Ni-NTA	

Protéine	Souche	Plasmide	Condition induction	Lyse	Batch/Co	onne Ni SEC/IEX
PII1-H2H3+SS4∆403	BL21 (DE3) pET300+ pETDuet	0,5 mM IPTG 20°C ON	l emul	siflex	voel 1
PII1-H2H3+SS4∆403	BL21 (DE3) pET300+ pETDuet	0,5 mM IPTG 20°C ON	l emul	siflex ml His Trap E	sec 25 ml xcel
PII1-H2H3+SS4∆403	BL21 (DE3) pET300+ pETDuet	0,5 mM IPTG 20°C ON	l emul	siflex ml+ dess	IEX 1 ml
PII1-H2H3+SS4∆403	BL21 (DE3) pET300+ pETDuet	1 mM IPTG 20°C ON	emul	siflex	IEX 1 ml + Ni Col 1 ml SEC
PII1-H2H3+SS4∆403	BL21 (DE3) pET300+ pETDuet	0,5 mM IPTG 20°C ON	l emul	His Trap siflex Excel ml Batch Ni-	25 ml
PII1-H2H3+SS4∆403	BL21 (DE3) pET300+ pETDuet	0,5 mM IPTG 20°C OI	N Sonic	ation NTA	
Protéine	Souche	Plasmide	Condition induction	Lyse	Batch/Colonne Ni	SEC/IEX
PII1-H2H3+SS4∆466 PII1-H2H3+SS4∆466	BL21 (DE3) BL21 (DE3)	pET300+ pETDuet pET300+ pETDuet	0,5 mM IPTG 20°C ON 0,5 mM IPTG 20°C ON	emulsiflex Sonication	Batch Ni-NTA	

Tableau 11. Conditions d'expression et de purification utilisées pour les co-purifications entre les constructions tronquées de PII1 et SS4.

Protéine	Souche	Plasmide	Condition induction	Lyse	Batch/Colonne Ni
PII1+ SS4∆349	BL21 (DE3)	pET300+pETDuet	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA
PII1∆71+ SS4∆349	BL21 (DE3)	pET300+pETDuet	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA
PII1∆101+SS4∆349	BL21 (DE3)	pET300+pETDuet	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA
PII1-H1-H2H3+SS4∆349	BL21 (DE3)	pET300+pETDuet	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA

PII1+ SS4∆403	BL21 (DE3)	pET300+pETDuet	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA
PII1∆71+ SS4∆403	BL21 (DE3)	pET300+pETDuet	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA
PII1∆101+SS4∆403	BL21 (DE3)	pET300+pETDuet	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA
PII1-H1-H2H3+SS4∆403	BL21 (DE3)	pET300+pETDuet	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA

PII1+ SS4∆466	BL21 (DE3)	pET300+pETDuet	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA
PII1∆71+SS4∆466	BL21 (DE3)	pET300+pETDuet	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA
PII1∆101+SS4∆466	BL21 (DE3)	pET300+pETDuet	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA
PII1-H1-H2H3+SS4∆466	BL21 (DE3)	pET300+pETDuet	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA

Tableau 12. Conditions d'expression et de purification utilisées pour le double clonage PII1- SS4∆349. SEC= chromatographie d'exclusion stérique. IEX= chromatographie échange d'anion. Dess= dessalage.

Protéine	Souche	Plasmide	Condition induction	Lyse	Batch/Colonne Ni	SEC/IEX
PII1-SS4∆349	BL21 (DE3)	pETDuet	0,5 mM IPTG 3h 37°C	sonication	Batch Ni-NTA	
PII1-SS4∆349	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA	
PII1-SS4∆349	BL21 (DE3)	pETDuet	AI 3h 37 °C	sonication	Batch Ni-NTA	
PII1-SS4∆349	BL21 (DE3)	pETDuet	AI 20°C	sonication	Batch Ni-NTA	
						Superdex
PII1-SS4∆349	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	Emulsiflex	His Trap Excel 1 ml	200
PII1-SS4∆349	Rosetta	pETDuet	0,5 mM IPTG 20°C ON	Emulsiflex	Batch Ni-NTA	
PII1-SS4∆349	Rosetta	pETDuet	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA + dess	IEX 1 ml
PII1-SS4∆349	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	sonication	batch Ni-NTA+ dilution	IEX 1 ml
PII1-SS4∆349	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	sonication	Batch Ni-NTA	SEC 25ml
PII1-SS4∆349	BL21 (DE3)	pETDuet	0,5 mM IPTG 20°C ON	Emulsiflex	Batch Ni-NTA	SEC 25ml

Chapitre 4. Analyse structurale

En parallèle des approches biochimiques visant à la production de PII1 et SS4 seules ou en complexe pour leur étude structurale *et du fait des difficultés rencontrées lors de ces étapes, j'ai entrepris leur étude structurale par des approches de bio-informatiques. Avec la révolution créee par AF dans ce domaine, j'ai décidé d'utiliser cet outil pour l'obtention de modèles moléculaires bien que la présence de CC prédits dans les deux protéines ne soit pas facile à prédire.*

Au début de ma thèse, AF2 (Jumper et al., 2021) venait d'être disponible notamment via la base de données Uniprot qui donnait accès à un modèle calculé pour certaines protéines dont PII1. Le modèle disponible n'ayant pas un indice de fiabilité très élevé et l'organisation en CC n'était pas prise en compte. Pour aller plus loin dans cette étude j'ai travaillé avec Dr. Guillaume Brysbaert de l'équipe de Dr. Marc Lensink qui développe le logiciel MassivFold (Brysbaert et al., 2024). MassivFold combine AlphaFold-Massive et CollabFold pour calculer de très nombreux modèles augmentant ainsi les chances d'obtenir un modèle fiable. Cette approche m'a permis d'obtenir un modèle formé principalement d'une longue région CC bien plus précis pour PII1 qui a plus tard été confirmé et affiné avec AF3 (J Abramson et al., 2024). Ensuite j'ai comparé la séquence et le modèle moléculaire de AtPII1 avec ceux de ses orthologues afin d'identifier des régions conservées. Enfin, nous avons confirmé la composition en structures secondaires d'AtPII1 par dichroïsme circulaire.

Dans un deuxième temps, dans l'optique d'identifier les régions d'interaction entre PII1 et SS4 et résidus impliqués, nous avons entrepris l'étude du complexe avec les mêmes approches. L'ensemble de ces résultats, associés aux résultats précédents, permet d'obtenir une meilleure compréhension de la dynamique structurale du complexe PII1/SS4.

4.1 Analyses bio-informatiques de la séquence de PII1

4.1.1 Analyse des propriétés physico-chimiques

Dans un premier temps, j'ai analysé les propriétés physico-chimiques de PII1 en soumettant sa séquence dans ProtParam (Gasteiger et al., 2005) qui fournit des données relatives à la séquence comme la composition en acides aminés, l'indice de stabilité, le pI théorique et le poids moléculaire ainsi que la moyenne générale d'hydropathie (GRAVY), l'indice aliphatique et le nombre de résidus chargés positivement et négativement. L'ensemble des données est récapitulé dans le **Tableau 13**.

Tableau 13. Caractéristiques physico-chimiques et composition en acide aminés de PII1. Les valeurs sont

Caractéristiques physico-chimiques				
Nombre d'acide aminés	783			
Poids moléculaire	90.30 Da			
pl théorique	4.5			
Index d'instabilité	53			
L' indice aliphatique	94.05			
Moyenne générale d'hydropathie (GRAVY)	-0.650			
Résidus chargés négativement (D, E)	160 (20.43%)			
Résidus chargés positivement (K, R, H)	130 (16.60%)			
Résidus polaires (C, S, Q, N, T, Y)	179 (22.86%)			
Résidus hydrophobes (A, G, I, L, M, P, F, W, V)	314 (40.10%)			

calculées par ProtParam à partir de la séquence de PII1 (Q8H1E5) (Gasteiger et al., 2005). Caractéristiques physico-chimiques Composition en acide aminés

Acide aminé	Nombre	% Total
Alanine (A)	58	7.41
Arginine (R)	47	6.00
Asparagine (N)	35	4.47
Acide aspartique (D)	31	3.96
Cysteine (C)	4	0.51
Glutamine (Q)	40	5.11
Acide glutamique (E)	129	16.48
Glycine (G)	19	2.43
Histidine (H)	15	1.92
Isoleucine (I)	27	3.45
Leucine (L)	112	14.30
Lysine (K)	68	8.68
Méthionine (M)	26	3.32
Phénylalanine (F)	13	1.66
Proline (P)	10	1.28
Serine (S)	57	7.28
Thréonine (T)	34	4.34
Tryptophane (W)	2	0.26
Tyrosine (Y)	9	1.15
Valine (V)	47	6.00

PII1 est une protéine très chargée avec 20% de résidus chargés négativement (D, E) et 16.6% de résidus chargés positivement (K, R, H). PII1 possède également une forte

proportion de leucines (14%). En revanche on remarque seulement deux tryptophanes, ce qui explique également pourquoi, les valeurs d'absorbance en mAU ne sont pas élevées lors de la purification. L'indice d'instabilité fournit une estimation de la stabilité de la protéine *in vitro*. Un indice supérieur à 40 signifie que la protéine est plutôt instable. Avec un indice de 53, PII1 est donc prédite instable, ce qui a été largement confirmé par nos essais d'expression et de purification.

L'indice aliphatique renseigne sur la thermostabilité des protéines globulaires (ikai, 1980). Il est défini comme le volume occupé par les chaînes latérales aliphatiques (alanine, leucine, isoleucine et valine). L'indice aliphatique de PII1 est de 93 indiquant que la protéine est thermostable sur une large gamme de température.

Enfin la valeur GRAVY renseigne sur la nature hydrophile ou hydrophobe de la protéine. PII1 est plutôt hydrophile et pouvant exercer une interaction forte avec les molécules d'eau.

4.1.2 Prédiction des éléments de structure secondaire

Pour connaitre la composition en structure secondaire de PII1, j'ai soumis la séquence à PSIPRED (McGuffin et al., 2000) et Gor4 (Garnier et al., 1996) dans le but de les comparer au modèle AF2 (**Figure 41**).



Figure 41. Prédiction in silico des éléments de structure secondaire à partir de la structure primaire de PII1 d'Arabidopsis thaliana avec (A) PSIPRED et (B) Gor4.

Les résultats indiquent la présence presque exclusive d'hélice a sur toute la longueur de la chaîne ainsi que deux potentiels feuillets β de prédits sur 3-4 résidus aux extrémités N- terminales et C-terminales. Ces résultats sont plutôt cohérents avec le modèle AF2 qui prédit une structure entièrement organisée en hélices.

En revanche, contrairement aux prédictions de CC qui prédisaient la présence de CC sur toute la longueur (Vandromme et al., 2019), PSIPRED et Gor4 prédisent des CC seulement aux extrémités N-terminales (1 à 105 résidus) et C-terminales (570 à 783 résidus) localisés dans les hélices H4, H5 et H6. Afin de vérifier plus exactement cette supposition nous avons réalisé d'autres prédictions de CC à partir de sites spécifiques disponibles.

4.1.3 Prédiction de coiled-coils

Dans le but d'affirmer ou d'infirmer la présence de CC dans la structure de PII1 nous avons utilisé le prédicteur de CC MarCoil (**Figure 42**), DeepCoils et PCOILS (Delorenzi et al., 2002, Ludwiczak et al., 2019, Gruber et al., 2006).



Figure 42. Prédiction de coiled-coils à partir de la structure primaire de PII1 d'Arabidopsis thaliana avec MarCoil. Les seuils sont tracés en ligne pleine horizontale, rouge pour les coiled-coils et bleu pour les hélices simples (SAH).

Les résultats du prédicteur MarCoil montrent la présence de CC prédits entre 100 et 570 résidus avec 100% de certitude c'est-à-dire entre H2 et H3. MarCoil prédit également des CC entre 600 et 730 résidus ce qui correspond aux hélices H4 et H5. En comparaison aux résultats précédents, le prédicteur certifie à 100% la présence de CC sur quasiment toute la structure de PII1. Ces résultats se répètent également chez d'autres prédicteurs de CC comme DeepCoil2 ou PCOILS.

4.1.4 Prédictions structurales de PII1

En premier lieu nous avons étudié la structure de PII1 à l'aide de AF2 (Jumper et al 2021) puis comparé cette structure à AF3 (Abramson et al., 2024) et à MassivFold (Brysbaert et al., 2024) (calculé au laboratoire) (**Figure 43**).

La confiance globale dans les prédictions pour les multimères doit prendre en compte tous les paramètres, pTM, ipTM, pLDDT et PAE.

Tout au long de ma thèse j'ai pu progressivement améliorer le modèle de PII1 en fonction de la disponibilité de nouveaux outils ou de nouvelles versions d'AF. Le premier modèle que j'ai analysé est celui disponible sur Uniprot (numéro d'accès) (**Figure 43A**). En effet les structures des différentes protéines d'organismes séquencés ont été calculées avec AF et rendues disponibles.



Figure 43. Modèles de PII1 d'Arabidopsis thaliana calculés par (A) AlphaFold2 et (B,C,D) MassivFold. En haut sont représentés les modèles moléculaires représentés en ruban et calculés en fonction de l'indice de fiabilité pLDDT. H= hélices. En bas sont représentés les matrices PAE pour chaque modèle. (D) Représentation des charges négatives (rouge) et positives (bleu) par résidus à la surface de la molécule (Jumper et al., 2021 ; Brysbaert et al., 2024).

Ce modèle (**Figure 43A**), présente une structure constituée uniquement d'hélices et de régions désordonnées. La structure est composée 6 hélices de longueurs variables et numérotées H1 à H6 à partir de l'extrémité N-terminale présentant un indice de confiance d'élevé à très élevé (**Figure 43A**). Les hélices H2 et H3 sont de très longues hélices comprenant 203 et 245 acides aminés respectivement et sont organisées en hélices simples. Les 70 acides aminés à l'extrémité N-terminale et la région entre les hélices H3 et H4 sont prédites désordonnées (pLDDT < 50). Enfin, la structure se complète avec trois autres hélices, H4, H5 et H6, de longueur plus petite et qui constituent la région C- terminale.

La matrice PAE indique que la position des hélices entre elles n'est pas fiable à l'exception des courtes hélices H5 et H6 entre elles. Ce modèle et notamment la prédiction des hélices nous a servi de base, faute de mieux, pour la réalisation de constructions tronquées mais ne présente pas un indice de confiance suffisant en termes de structure globale (cf. **Chapitre 1**). Il ne montre notamment pas l'organisation des hélices en CC qui est prédite par des outils de prédictions et qui a un sens biologique si on considère que les

partenaires de PII1 *in vivo* ont également des régions CC prédites qui expliqueraient leur mode d'interaction.

Afin d'améliorer la qualité du modèle je me suis rapprochée de Dr. Guillaume Brysbaert dans l'équipe de Dr. Marc Lensink au laboratoire qui développe des outils visant à optimiser la modélisation moléculaire avec AF. MassivFold permet d'accéder à une diversité structurelle des modèles pour les structures. Il offre la possibilité de calculer cinq modèles pour les monomères et quinze pour les multimères et augmente les possibilités d'obtenir un modèle fiable. Sur la base de la séquence j'ai calculé cinq modèles de PII1 et obtenu le résultat **Figure 43B**.

La structure est modélisée entièrement en hélice et organisée de la même façon que le modèle calculé par AF2. Le modèle prédit cependant une organisation en CC des hélices H2 et H3. On aperçoit également, une interaction probable entre H1 et H2 et H5 et H6. Ce modèle est a priori plus proche de ce qui a été décrit précédemment, car il comporte les hélices H2 et H3 organisées en CC. En revanche la matrice PAE ne semble pas fiable dans le sens où elle ne valide pas l'interaction des hélices H2 et H3.

Afin de vérifier et affiner l'organisation en CC de la structure nous avons modélisé uniquement les hélices H2 et H3 d'une part car elles constituent la construction PII1-H2H3 utilisées pour les expériences mais également pour ne garder que les régions repliées. En effet les régions non repliées pourraient gêner le score global du modèle (**Figure 43C**).

L'analyse montre que les hélices H2 et H3 sont modélisées en CC avec un indice de confiance plDDT supérieur à 90 et valide les résultats obtenus en prédiction de CC. De plus, la matrice PAE valide le positionnement de ces hélices.

Les CC présents dans la structure de PII1 confortent le fait qu'elle soit capable d'interagir avec des partenaires. Pour cela la répartition des charges par résidus a été calculé avec PYMOL. On observe que PII1 est très chargée à la surface des hélices, ce qui ce qui peut faciliter les interactions avec le solvant ou des protéines.

A la sortie d'AF3 en 2024, j'ai resoumis la séquence de PII1 et PII1-H2H3 afin de les comparer à nos modèles MassivFold (**Figure 44**).

133



Figure 44. Modèles de (A) PII1 d'Arabidopsis thaliana et (B,C) PII1-H2H3 calculés par AlphaFold3. En haut sont représentés les modèles moléculaires représentés en ruban et calculés en fonction de l'indice de fiabilité pLDDT. H= hélices. En bas sont représentés les matrices PAE pour chaque modèle. (C) Représentation des charges négatives (rouge) et positives (bleu) par résidus.

La structure prédite de PII1 et PII1-H2H3 par AF3 est identique à celle obtenue précédemment avec MassivFold. Le score pLDDT est supérieur à 90 pour les hélices H2 et H3 et celles-ci sont repliées en CC mais contrairement à MassivFold, la matrice PAE valide cet agencement même pour PII1 entière. Etant donné que nous avons obtenus de meilleurs résultats avec AF3, nous l'avons donc utilisé pour la suite des analyses.

En résumé, à partir de tous ces modèles, nous avons un modèle fiable de la structure de PII1 et particulièrement de ces deux hélices qui sont modélisées en CC. Afin de vérifier si la structure globale prédite est conservée chez les autres espèces et pour voir si les deux hélices H2 et H3 sont également en cc, nous avons débuté une étude sur les orthologues de PII1.

4.1.6 Analyse de l'alignement de séquence avec les orthologues

J'ai entrepris une analyse bio-informatique avec orthologues de PII1. Pour cela j'ai extrait la séquence des 28 orthologues disponibles sur le site The Arabidopsis Information Resource (TAIR) (https://www.arabidopsis.org/] (**Tableau 14**). Ce site récence tous les gènes issus de la plante supérieure *Arabidopsis thaliana* et renseigne également sur les orthologues potentiels d'une protéine donnée. Les orthologues utilisés pour l'alignement de séquence, le nombres de résidus, l'identité de séquence et le recouvrement sont résumés dans le **Tableau 14**.

N°	Organisme	Nom	Longueur séquence (aas)	ldentité de sequence (%)	Recouvrement (%)
1	Arabidopsis thaliana		783	100	100
2	Gossypium hirsutum	Coton mexicain	778	53	98
з	Prunus persica	Péchier	788	53	99
4	Ricinus communis	Ricin commun	776	52	98
5	Citrus sinensis	Oranger	771	52	98
6	Juglans regia	Noyer	771	53	98
7	Nelumbo nucifera	Lotus sacré	784	51	98
8	Manihot esculenta	Manioc	773	52	98
9	Eucalyptus grandis	Gommier rose	776	52	98
10	Populus trichocarpa	Peuplier de l'ouest	771	52	98
11	Theobroma cacao	Cacao	817	49	98
12	Vitis vinifera	Raisin	774	51	97
13	Nicotiana tabacum	Tabac	766	49	97
14	Cucumis sativus	Concombre	777	48	97
15	Erythranthe guttata	Mimule jaune	774	47	97
16	Solanum tuberosum	Pomme de terre	770	46	97
17	Glycine max	Soja	718	48	91
18	Solanum lycopersicum	Tomate	770	46	97
19	Medicago truncatula	Luzerne tronquée	755	45	98
20	Capsicum annuum	Poivre	730	46	91
21	Helianthus annuus	Tournesol	748	43	98
22	Amborella trichopoda		800	38	96
23	Lactuca sativa	Laitue	729	38	98
24	Brachypodium distachyon	Brachypode à deux épis	780	34	91
25	Sorghum bicolor	Sorgo	774	32	91
26	Zea mays	Mais	771	33	91
27	Oryza sativa	Riz	763	34	96
28	Triticum aestivum	Blé	773	33	91
29	Setaria italica	Panis	492	34	60

Tableau 14. Orthologues de PII1 d'Arabidopsis thaliana utilisés pour l'alignement de séquence multiple.

À partir des séquences disponibles, j'ai réalisé un alignement multiple avec COBALT (Papadopoulos et al., 2007). Un aperçu graphique est disponible en (**Figure 45**) et l'alignement entier est disponible en (**Annexe 4**).



Figure 45. Analyse des régions et résidus conservées chez PII1 d'Arabidopsis et les orthologues. (A) Alignement de la séquence d'acides aminés de PII1 avec les 27 orthologues de la même longueur calculée avec Blastp COBALT (Papadopoulos et al 2007). La séquence de Setaria italica a été ajoutée manuellement à partir de l'alignement de PII1 avec tous les orthologues testés. Les résidus conservés sont colorés en rouge et les résidus moins conservés sont colorés en bleu. Les encadrés jaune et vert correspondent aux hélices H1 et H2/H3 respectivement selon le modèle calculé en (B). (B) Modèle AlphaFold3 de PII1en ruban coloré en fonction de la conservation des résidus identifiés par ConSurf. L'échelle de conservation varie du bleu au rouge en fonction des résidus d'une conservation variable (bleu) aux résidus conservés (rouge). Les résidus conservés identifiés par Blastp sont représentés en stick sur le modèle et listé dans le tableau de droite.

Tous les orthologues sont de taille proche, de *At*PII1, entre 718 et 817 résidus, excepté chez la plante *Setaria italica* (*Si*) dont la séquence est beaucoup plus courte (492 aa). Leur séquence a été alignées avec Blastp. Les résultats montrent une identité de séquence qui varie entre 53% et 33% selon les espèces végétales et donc très conservée chez tous les orthologues, tout le long de la séquence. Ces résultats indiquent que toutes les hélices sont

conservées. De plus on observe que *Setaria* possède une séquence beaucoup plus courte que les autres orthologues correspondant uniquement à la région N-terminale contenant H1, H2 et H3 dans *At*PII1 (Encadré vert et jaune **Figure 45B**). Ces résultats montrent que H1, H2 et H3 sont conservées chez toutes les protéines identifiées mais les H4, H5 et H6 sont également conservée. L'alignement a permis de mettre en évidence 21 résidus très conservés dont 13 leucines situés dans l'hélice H2 et H3 (**Figure 45B**). Ces résidus conservés sont surlignés en rouge dans l'**Annexe 4**.

4.1.7 Analyse de la structure des orthologues avec AlphaFold3

Comme pour *At*PII1, nous avons calculé un modèle moléculaire pour chacun des orthologues PII1 identifiés dans TAIR à l'aide d'AF3 (**Figure 46**). Par souci de clarté, j'ai représenté uniquement les hélices H2 et H3.

	Organisme	RMSD
	Gossypium hirsutum	2.927
	Prunus persica	3.081
	Ricinus communis	2.939
	Citrus sinensis	3.606
	Juglans regia	2.681
	Nelumbo nucifera	3.929
• H3	Manihot esculenta	3.823
	Eucalyptus grandis	3.443
	Populus trichocarpa	4.343
	Theobroma cacao	3.094
Transferrer and the second state of the second	Vitis vinifera	4.770
	Nicotiana tabacum	3.459
	Cucumis sativus	3.282
and the second s	Erythranthe guttata	3.871
	Solanum tuberosum	2.821
	Glycine max	3.443
	Solanum lycopersicum	3.492
	Medicago truncatula	4.576
	Capsicum annuum	3.543
	Helianthus annuus	3.695
	Amborella trichopoda	3.577
	Lactuca sativa	2.265
	Brachypodium distachyon	1.859
	Sorghum bicolor	3.284
	Zea mays	4.707
	Oryza sativa	2.231
	Triticum aestivum	3.069
	Setaria italica	4.794

Figure 46. Superposition des modèles AlphFold3 de la structure hélices H2 et H3 de PII1 avec ceux des 28 orthologues calculés par AlphaFol3.Les structures ont été superposées avec le logiciel PyMOL. Les hélices et les boucles sont colorées en rouge et en vert respectivement. Les valeurs de RMSD de l'alignement entre PII1 d'Arabidopsis et les 28 orthologues sont indiqués dans le tableau de droite.

A l'exception de *Setaria* qui ne contient que les hélices H1, H2 et H3, tous les autres orthologues montrent la présence de six hélices structurellement similaires à la structure de PII1 d'*Arabidopsis*. Ces deux hélices H2 et H3 sont modélisées en CC anti-parallèles et les hélices H2 et H3 sont quasiment de la même longueur chez tous les orthologues (**Figure 46**). L'alignement des structures de H2 et H3 entre *At*PII1 et les autres orthologues a été calculée. Les résultats montrent un degré élevé de conservation structurelle de la région CC, ce qui est en accord avec le rôle important de cette région dans la fonction de PII1. Le RMSD a été utilisé pour comparer les superpositions de structures de protéines et évaluer la qualité des alignements entre *At*PII1 et les 28 orthologues. Les valeurs de RMSD des hélices H2 et H3 sont comprises entre 1.8 et 4.9 indiquant que l'alignement est plutôt bon. L'écart entre les valeurs de RMSD peut s'expliquer par le fait que la distance spatiale entre les atomes est plus grande aux extrémités des hélices H2 et H3 ce qui est cohérent avec le fait que les protéines hélices H2 et H3 associées en CC peuvent adopter une structure relativement flexible.

4.1.8 Analyse structurale de PII1-H2H3 par dichroïsme circulaire par rayonnement synchrotron

Analyse du contenu en structure secondaire de PII1-H2H3 à 25°C

Pour vérifier les prédictions de structure et valider la présence d'hélices prédites dans PII1-H2H3, nous avons réalisé des expériences de SR-CD sur la ligne de lumière DISCO du synchrotron SOLEIL. Pour cela nous avons réalisé plusieurs expériences à partir des échantillons protéiques PII1-H2H3 purifiés par affinité et SEC. La protéine a été concentrée à 2,5 mg/ml. Premièrement, nous avons calculé le contenu en structure secondaire de l'échantillon à partir d'une mesure de spectre réalisée à 25°C. Nous avons collecté le spectre CD sur une gamme de longueurs d'onde allant de 170 à 270 nm. La déconvolution du spectre qui sert à déterminer le contenu en structure secondaire a été réalisée avec BestSel. Les résultats sont présentés en **Figure 47**.



Figure 47. Purification et analyse SR-CD de PII1-H2H3. (A) SDS-PAGE de l'échantillon protéique PII1-H2H3 après purification, utilisé pour les expériences SR-CD. (*B*) Spectre SR-CD de PII1-H2H3.

L'analyse SR-CD présente un profil typique d'une protéine repliée en hélice avec un pic positif à 190nm et deux pics négatifs à 208nm et 222 nm. La déconvolution du spectre a permis de trouver un contenu en hélice supérieur à 98% (NRMSD of 0,021) avec 0,9% de feuillets β anti parallèle à 25°C. Ce résultat, bien que probablement un peu surestimé, renforce la fiabilité de notre modèle montrant un contenu en structure secondaire exclusivement composée d'hélices a. Avec cette approche, il est également possible de distinguer la présence d'hélices simples de la présence d'hélices impliquées dans des CC en calculant le ratio des pics à 222/208 nm (Minelli et al., 2013). Celui-ci être supérieur ou égal à 1,0 (+/- 0,03) pour signifier que les hélices sont agencées en CC. À 25°C le ratio calculé est de 1,07, ce qui indique et confirme l'organisation des hélices en CC.

Effet du TFE sur PII1-H2H3

Pour confirmer la présence de CC nous avons ajouté dans le tampon de la protéine, des quantités croissantes de TFE (10%, 20%, 30%, 50% et 60%). Nous avons utilisé la même quantité de protéines dans chaque échantillon en ajustant le volume final avec le tampon. Le TFE permet de créer un environnement hydrophobe qui masque les régions qui servent d'interaction entre les hélices ou avec des partenaires. De ce fait les hélices a individuelles seront stabilisées et les CC se dissocieront (Nelson and Kallenbach., 1986). Cette expérience nous confirmera la présence ou non de CC chez PII1-H2H3 (**Figure 48**).



Figure 48. Influence de l'augmentation de la quantité de TFE sur PII1-H2H3 par SR-CD. Les mesures ont été effectuées à 25°C sur un échantillon protéique concentré à 2,5 mg/ml.

Les résultats montrent un effet plutôt atypique du TFE sur PII1-H2H3 à de faibles concentrations. A 10% de TFE, on observe un aplatissement du spectre. Le pic à 222 nm perd en intensité ce qui signifie que les hélices de la protéine sont initialement dépliées puis

en rajoutant 20 ou 30% de TFE, les hélices se reforment progressivement mais perdent leur association en CC (Θ 222/208 nm < 1) (**Figure 48**). En augmentant la concentration à 50% ou 60% de TFE, on retrouve un signal maximal correspondant à une restructuration de la protéine dans laquelle les hélices a sont indépendantes uniquement. Ces résultats nous confirment que la structure de PII1-H2H3 possède des hélices en CC. Le fait que PII1-H2H3 puisse se déplier facilement puis reformer des hélices est tout à fait inattendu et montre que la protéine possède une structure flexible et adaptable dans certaines conditions, ce qui lui permet de subir des changements conformationnels majeurs. Par conséquent, il est probable qu'elle puisse utiliser cette caractéristique pour interagir avec elle-même ou avec d'autres partenaires de l'initiation.

4.2 Prédictions structurales de SS4, SS5 et des complexes PII1/SS4 et PII1/SS5

4.2.1 Analyse de la structure de SS4 dans une forme monomérique

Quatre régions CC numérotées de CC1 à CC4 et une région de dimérisation RD avaient été prédites par (Raynaud et al., 2016). Dans un premier temps nous avons modélisé la structure de SS4 sous forme monomérique avec différents prédicteurs de structure, AF2, MassivFold et AF3 (**Figure 49**).



Figure 49. Modèles de SS4 calculés par (A) AlphaFold2, (B) MassivFold et (C) AlphaFold3. Les structures sont représentées en ruban. Les régions sont colorées selon le niveau de confiance du modèle (pLDDT). Les matrices PAE sont indiquées sous chaque modèle. H = Hélice. RD= région de dimérisation. GT= domaine glycosyltransférase.

Les modèles prédits par AF2, AF3 et MassivFold montre d'une part une région N- terminale composée d'une région déstructurée et des hélices dont le nombre varie selon le prédicteur de structure et d'autre part d'un domaine catalytique qui contient le GT1 et le GT5. Entre les deux se trouve une région RD constituée de 4 petites hélices qui correspond à la région de dimérisation décrite par (Raynaud et al., 2016). La région N-terminale et la RD sont prédites avec un bon indice de confiance (pLDDT entre 70 et 90) et le domaine catalytique est prédit avec un indice de confiance très élevé (pLDDT supérieur à 90) et bien conservé dans les trois modèles.

Le modèle AF2 est décrit avec quatre hélices tandis que le modèle MassivFold et AF3 sont prédits avec 3 ou 2 hélices plus grandes respectivement. En comparaison avec les prédictions de (Raynaud et al., 2016), qui décrivait les hélices comme étant impliquées dans quatre régions CC, ici les hélices sont individuelles. L'analyse de la matrice PAE montre un mauvais niveau de confiance sur la position de ces hélices. Le domaine catalytique est prédit avec un niveau de confiance élevé (pLDDT supérieur à 90) et superposable à la structure résolue par cristallographie aux rayons X (RMSD à 0,3 Å) **Figure 50**.



Figure 50. Superposition de la structure du domaine catalytique (GT1 et GT5) de SS4 résolue par cristallographie (en vert) avec le domaine catalytique et la région de dimérisation (RD) de SS4 calculée par AlphaFold3 (en bleu). Le site catalytique est indiqué par une étoile. L'acarbose et le glucose sont colorés en orange. GT= glycosyltransferase. Le domaine de dimérisation est encadré.

Le domaine de dimérisation tel que décrit par Raynaud et al., (2016), est situé entre les résidus 471 et 515 (Raynaud et al., 2016). Le modèle AF3 montre que le RD se situe entre résidus 473 et 534. La structure est globulaire et composée de quatre hélices (H5 à H8). Ces hélices sont parallèles entre elles, deux par deux, et prédites avec un bon niveau de confiance (pLDDT entre 70 et 90) (**Figure 50**).

A partir de ces résultats, nous nous sommes demandé si SS5, décrite également comme étant une protéine dimérique, possède une région de dimérisation structurellement similaire à SS4.

4.2.2 Alignement de la région de dimérisation entre SS4 et SS5

SS5 est une protéine impliquée dans les mécanismes d'initiation, interagissant avec PII1 et décrite comme étant une protéine dimérique (Abt et al., 2020). De plus, SS5 comporte seulement le domaine GT5 localisé en position C-terminale contrairement aux autres SSs qui possède le GT5 et le GT1. Nous avons modélisé la structure de SS5 avec AF3 et aligné ce modèle avec la région catalytique et la région RD de SS4 (**Figure 51**).



Figure 51. Alignement de structure entre SS4 et SS5. (A) Modèle AlphaFold3 de SS5 (la matrice PAE est représentée en bas) superposé à la structure du domaine catalytique de SS4 (pdb code 6GNE) en orange. (*B*) Zoom sur le domaine catalytique et RD de SS4 (orange) superposés avec ceux de SS5 (bleu). (*C*) Zoom sur le domaine RD de SS4 (orange) superposé avec celui de SS5 (bleu). Les structures sont représentées en ruban et l'alignement est réalisé avec PyMOL. RD= région de dimérisation. GT= domaine glycosyltransférase.

Le modèle obtenu montre une hélice entre les résidus 60 et 139 puis trois autres plus petites et prédites avec un bon niveau de confiance (pLDDT entre 70 et 90). Le domaine GT5 est prédit avec un très haut niveau de confiance (pLDDT > à 90). De plus, ce domaine GT5 se prolonge par une hélice 439-460 située à l'extrémité C-terminale, rattachée et interagissant avec le GT5. Cette hélice, de SS5, est structurellement semblable à l'hélice
C- terminale 1015-1035 de SS4, rattachée au GT1 et interagissant avec le GT5 et prédite avec un bon indice de confiance.

L'alignement de la structure de SS5 et SS4 montre que ces quatre hélices correspondent à la région de dimérisation de SS5 et sont superposables. La région de dimérisation RD, le domaine GT5 et l'hélice 439 -460 de SS5 se superposent aux structures équivalentes de SS4 avec un RMSD de 1 Å. SS4 et SS5 sont les seules SS dimériques (et possèdent un domaine RD) et interagissent avec PII1 donc il est possible que la région de dimérisation soit impliquée dans cette interaction. L'interaction entre PII1 et SS4 est décrite dans le paragraphe 4.2.4.

4.2.3 Analyse de la structure de SS4, SS4∆349 et SS5 sous forme dimérique

Les travaux de Raynaud et al.,2016, ont permis de caractériser SS4 comme étant une protéine dimérique, dépendante de la région RD. La structure de cette région N-terminale n'a pas été déterminée en revanche la région C-terminale constituée du domaine catalytique a été résolue par cristallographie (Nielsen et al., 2018). Nous avons réalisé des modélisations avec AF3 de SS4 et S55 en dimère pour améliorer la qualité des modèles et des matrices PAE qui ne sont pas optimales pour les modèles des monomères (**Figure 52 A, C**).

Enfin, lors de ces travaux plusieurs constructions tronquées de SS4 ont été réalisées et la suppression de CC1 et CC2 (délétion des 349 premiers résidus) a permis d'obtenir un échantillon protéique soluble (Raynaud et al., 2016). De ce fait, en parallèle de la protéine SS4 entière, nous avons réalisé des modèles sur les constructions SS4∆349 en dimère (**Figure 52 B**).



Figure 52. Modèles AlphaFold3 en dimère de (A) SS4, (B) SS4∆349 et (C) SS5. Les modèles de gauche sont représentés en ruban et les régions sont colorées selon l'indice de confiance du modèle (pLDDT). Le modèle est accompagné de la matrice d'erreur PEA. Les modèles centraux représentent l'organisation structurale des protéines. Seul un monomère est coloré (hélices de la région N-terminale en bleu, RD en orange, domaine catalytique en vert et hélice C-terminale en rouge). Le deuxième monomère est coloré en gris, ruban et en surface moléculaire avec pyMOL. Le modèle de droite représente le domaine catalytique, la RD et l'hélice C-terminale, représenté avec le même code couleur que le modèle central.

L'analyse des modèles SS4, SS5 et SS4 Δ 349 montre les domaines catalytiques prédits avec un niveau de confiance élevé (pLDDT supérieur à 90) et la région RD avec un bon niveau de confiance (pLDDT entre 70 et 90) avec un score ipTM et pTM pour tous les modèles de 0,4 à 0,5.

Les hélices N-terminales sont modélisées différemment pour SS4Δ349 par rapport à SS4 entière. En effet on observe deux petites hélices (H3 et H4) correspondant à la grande hélice H2 de SS4. Dans ce modèle les hélices H3 et H4 interagissent ensemble en CC, avec un bon niveau de confiance (pLDDT entre 70 et 90) et une position correcte des hélices entre elles visible sur la matrice PAE. Il est probable que AF3 ait positionné ces CC pour stabiliser le modèle et d'après la matrice PAE, il est possible que les régions CC prédites s'organisent vraisemblablement en CC intermoléculaires et non pas intramoléculaires. Comme pour la protéine SS4 entière, cela pourrait indiquer que les régions CC individuelles sont capables d'adapter leur repliement en fonction d'un partenaire ou plusieurs partenaires individuellement ou en même temps.

Chaque modèle des protéines dimériques montre que la région RD d'un monomère interagit avec la sous-unité GT5 du second monomère impliquant des interactions hydrophiles et hydrophobes. L'analyse des matrices PAE confirme cette interaction entre le RD et le GT5. Cet agencement est retrouvé chez SS4, SS5 et SS4 Δ 349 ce qui crédibilise la fiabilité du modèle. De plus, l'interaction entre les RD observés dans le modèle SS4 ou SS5 n'est pas conservée dans le modèle SS4 Δ 349 dans lequel les deux régions de dimérisation ne sont pas dans la même position et n'interagissent pas directement. Dans le contexte de l'initiation de la biosynthèse, cela pourrait indiquer l'interaction entre le RD d'un monomère et le GT5 du second monomère resterait stable tandis que les deux RD des protéines dimériques pourrait subir un changement de conformation en fonction de l'association du ligand (MOS) avec SS4/SS5 ou de l'interaction avec un autre partenaire de l'initiation. Pour cela nous avons modéliser les complexes entre SS4/S55 en monomère et dimère avec PII1.

4.2.4 Prédiction structurale du complexe PII1 et SS4

De nombreuses interactions entre les protéines impliquées dans l'initiation ont été démontrées. Parmi elles, PII1 interagit avec SS4 et également avec SS5 (Seung et al., 2018 ; Abt et al., 2020). Nous avons précédemment décrit le modèle PII1 et PII1-H2H3 et SS4 (et SS5) en dimère affichant un bon niveau de confiance. Pour visualiser les modes d'interaction entre PII1 et SS4/SS5 nous avons modélisé des structures de complexes en utilisant différentes approches. Premièrement nous avons utilisé plusieurs logiciels de modélisation (MassivFold et AF3). Nous avons généré cinquante modèles avec AF3 et quinze modèles avec MassivFold de SS4 en complexe avec PII1 puis analysé les modèles. Malheureusement, nous n'avons pas pu obtenir de modèle fiable. Les scores iPTM et pTM très bas (inférieur à 0.3) ainsi que le placement des molécules l'une par rapport à l'autre n'était pas valide, comme le montrait la matrice PAE. Les valeurs iPTM et pTM de tous les modèles calculés n'ont jamais dépassé 0,4.

Nous avons affiné les modèles en utilisant soit la construction tronquée de SS4Δ349 en dimère ou en monomère en complexe avec PII1-H2H3. Nous avons représenté le résultat répondant à nos critères, c'est-à-dire le complexe PII1-H2H3/SS4Δ349 par MassivFold (**Figure 53**).



Figure 53. Modèle de structure du complexe entre SS4 SS4Δ349 (vert) et PII1-H2H3 (magenta) par MassivFold. La structure est représentée en ruban et accompagnée de la matrice PAE.

Dans ce modèle, le score global est bon pour la modélisation des hélices de PII1- H2H3 et SS4Δ349 (pLDDT entre 70 et 90) et très bon pour le domaine catalytique de SS4Δ349 (pLDDT supérieur à 90). PII1-H2H3 est modélisée avec deux hélices en CC qui interagissent avec la région de dimérisation et l'hélice de SS4 comme le confirme la matrice PAE. Le score pTM et iPTM du complexe est de 0,4.

Lors de nos modélisations, nous avons observé que le prédicteur de structure utilisé ou les constructions protéiques choisies pouvaient changer drastiquement la structure des protéines modélisées et la région d'interaction. Par exemple les hélices de PII1 ou PII1-H2H3 n'étaient pas toujours modélisées en CC alors que nous l'avons démontré expérimentalement. La même chose a été observée pour SS4 ou les constructions tronquées lorsqu'elles étaient modélisées avec PII1. Ces changements de conformation très flexible peuvent faire penser que ces protéines puissent s'adapter à un ou plusieurs partenaires.

En conclusion au vu de la difficulté à modéliser les complexes étant donné les scores beaucoup plus faibles que la modélisation des protéines seules, nous ne pouvons pas être certain du positionnement de PII1 par rapport à SS4 ou SS5 cependant il semblerait que le complexe se forme au niveau de la région RD/ GT5. Du fait de la difficulté de purification des complexes, nous avons décidé de travailler sur des extraits moins purs des protéines seules pour des essais enzymatiques.

4.3 Prédictions structurales de PTST2, PTST3, SS3 et MFP1

A ce jour, la modélisation des structures des autres participants de l'initiation tels que PTST2 et PTST3, MFP1 ou encore SS3 (qui a été décrite comme étant impliquée dans l'initiation en absence de SS4 (Szydlowski et al., 2009)) n'ont pas été réalisées avec AF3. Dans ce travail j'ai modélisé avec AF3, la structure de ces protéines entières (**Figure 54**). Ces résultats donneront un aperçu de leur structure et nous permettrons de vérifier les prédictions des domaines réalisées, c'est-à-dire, la présence d'un CBM48 pour PTST2 et PTST3 et trois CBM53 pour SS3 et la présence de CC présents chez toutes ces protéines.



Figure 54. Modèles AlphaFold3 de PTST2 (A), PTST3 (B), MFP1 (C) et SS3 (D). (A), (B) et (C) : les structures sont modélisées en ruban et colorées selon le score de fiabilité du modèle avec le pLDDT (de 0 à 100). pLDDT > 90 régions colorées en bleu foncé, les régions entre 90 >pLDDT > 70 sont colorées en cyan, les régions avec pLDDT < 70 ou < 50 sont colorées en jaune et orange respectivement. (D) Structure SS3 modélisée en ruban et colorée selon les domaines (domaine catalytique GT1 et GT5 en vert, CBM en rose, autre en gris). Les modélisations sont accompagnées de la matrice PAE associée.

Les résultats des modèles prédits pour PTST2 et PTST3 montrent qu'un CBM est présent et modélisé avec un très bon score (pLDDT > 90) tout comme l'hélice qui le précède (Figure 54 A,B). L'alignement du CBM et de l'hélice est bon avec un RMSD à 1.8 Å. Le reste de leur structure n'est pas correctement définie (pLDDT < à 70, 50) avec de nombreuses régions non repliées, mis à part un petit domaine globulaire de quatre hélices (entouré en pointillé dans la Figure 54). Ce domaine, dont la fonction est inconnue, est superposable entre PTST3 et PTST2 avec un RMSD de 0.7 Å. Le modèle prédit pour MFP1 ressemble à ce qu'il est décrit dans la littérature c'est- à- dire sans domaine particulier. La structure est allongée et contient deux longues hélices avec un score pLDDT fiable (>90) (**Figure 54C**). MFP1 est décrite avec des régions CC sur toute sa longueur. D'après les cinq modèles calculés, les hélices étaient ou non associées en CC. On peut supposer que, comme PII1, ces longues hélices puissent s'associer quand même entre elles, d'autant plus que la matrice PEA de ce modèle n'est pas convaincante.

En parallèle j'ai modélisé la MFP1 en complexe avec PTST2, protéine avec laquelle elle interagit dans l'initiation (Sharma et al 2024), pour vérifier si le repliement des hélices de MFP1 peut changer avec la présence d'un partenaire (résultats non représentés car les scores ipTM et TM du complexe étaient < à 0.25). Dans ce cas, j'ai observé, que les hélices de MFP1 en complexe avec PTST2 étaient repliées en CC et la matrice PEA pour MFP1 seulement, en complexe était améliorée et montrait une interaction entre les deux longues hélices. Ce résultat signifie probablement que MFP1 est (comme PII1) une protéine dynamique et qui peut s'adapter à son environnement/ partenaire. Il est probable également qu'au vu de sa fonction d'encrage dans la membrane des thylakoïdes, le CC est peut-être rigidifié par la région transmembranaire de la région N-terminale.

Ensuite, j'ai modélisé la structure de SS3 étant donné qu'elle participe également à l'initiation en absence de SS4 (**Figure 54D**). Les résultats montrent la présence d'un domaine catalytique GT1 et GT5, semblable à toutes les SSs et trois domaines globulaires CBM48 impliqués pour l'interaction avec les glucanes et tous prédits avec un score de fiabilité très élevé (pLDDT >90). La matrice PAE de cette structure est particulièrement bonne puisque tous les domaines semblent être bien positionnés les uns avec les autres. Chaque domaine CBM est rattaché à une longue hélice forme une région qui rappelle la structure de PTST2 et avec laquelle elle serait capable d'imiter à la fois la fonction de SS4 et de PTST2 en leur absence (Szydlowski et al., 2009).

Enfin, étant donné que PTST2 interagit également avec SS4, PII1 et MFP1, j'ai modélisé également PTST2 en complexe avec SS4 dimérique ou PII1 ou MFP1. Cependant, bien que les scores pLDDT des structures étaient semblables aux modèles des protéines monomériques, les scores iPTM et pTM étaient très bas (entre 0,13 et 0,25) et la matrice PEA

151

montrait aucune interaction fiable entre ces molécules donc ces modèles ne sont pas présentés ici car ils ne respectent pas les critères de validité.

En conclusion, toutes les informations structurales réunies complètent les données de génétique inverse et des interactions, et peuvent fournir de nouvelles pistes pour l'analyse des rôles de chaque protéine dans le mécanisme d'initiation.

Chapitre 5. Etude enzymatique de PII1 et SS4

SS4 possède une activité enzymatique de glycosyltransférase et intervient pendant l'initiation (Roldán et al., 2007). A notre connaissance aucune étude n'a montré d'activité des protéines SS4 extraites de feuilles par (Szydlowski et al., 2009). Il a été proposé que l'absence d'activité détectable de SS4 sur les extraits végétaux soit due à sa présence en faible quantité dans les extraits. En produisant des protéines tronquées nous avons montré qu'il était possible d'obtenir des quantités beaucoup plus importantes des enzymes plus appropriées aux tests d'activité. Afin de vérifier que SS4 était active dans les formes tronquées (SS4D349, SS4D403 et SS4D466) nous avons réalisé des tests d'activité par zymogramme sur ces protéines. Ce test consiste à faire migrer les protéines dans un gel de polyacrylamide en conditions natives contenant du glycogène comme substrat. Après migration, les gels sont incubés avec de l'ADPG qui est le donneur de glucose de la réaction. La production de chaînes de polyglucanes par l'enzyme est ensuite révélée par une coloration à l'iode. Du fait de la difficulté avérée de purifier SS4 et pour diminuer le risque d'altération des enzymes, nous avons utilisé les extraits bactériens solubles plutôt que purifiés pour les zymogrammes. Le but de ces expériences était dans un premier temps de vérifier l'activité des constructions tronquées pour, dans un deuxième temps, pouvoir étudier un éventuel effet de la présence de PII1 sur leur activité.

5.1 L'activité catalytique d'élongation de SS4 est fonctionnelle

Dans un premier temps, nous avons réalisé des gels d'activités en gel natif de polyacrylamide 10% en présence glycogène (0,3%) pour étudier la capacité de SS4 à allonger les chaînes de glucanes. Le gel a ensuite été incubé une nuit dans un mélange contenant de l'ADPG comme donneur de glucose et les activités sont révélées après coloration à l'iode (**Figure 55**). Nous avons également incubé le gel dans le mélange sans ADPG pour servir de témoin contre l'action possible de l'enzyme disporportionante (D- enzyme) d'*E.coli* qui peut cliver les chaînes de glycogène qui seraient alors visibles par

révélation à l'iode. Le glycogène possède des chaînes courtes qui ne peuvent pas fixer l'iode. Au contraire si les glycosyltransférases catalysent l'allongement des chaînes d'amylopectine, l'iode se fixe entre les chaînes plus longues qui se traduit par une coloration brune.

Dans un premier temps, nous avons a déposé l'équivalent de 150 μ g d'extrait soluble pour les 3 formes tronquées de SS4 (SS4 Δ 349, SS4 Δ 403 et SS4 Δ 466). De la même façon, on a déposé la même quantité de l'extrait de souche BL21(DE) non transformée pour différencier l'activité de glgA, la glycogène synthase présente chez *E. coli* et de SS4 (**Figure 55**).



Figure 55. Caractérisation des activités d'élongation de SS4. Les extraits solubles des constructions tronquées SS4Δ349, SS4Δ403 et SS4 Δ466 ont été déposés sur gel natif 10% incubés avec (**A**) ou sans ADP-glucose (**B**). L'extrait soluble BL21(DE3) qui n'a subi aucune transformation a également été déposé pour identifier l'activité de la glycogène synthase glgA d'E.coli. Les flèches noires correspondent aux activités synthases de SS4.

L'analyse des résultats montre que les gels qui n'ont pas été incubés avec de l'ADPG ne montrent aucune activité que ce soit pour les constructions tronquées et la glgA. Ce qui est cohérent car sans le donneur de glucose, les SSs ne sont pas capables d'allonger les chaînes de glucanes. Ensuite l'analyse des résultats des gels incubé avec de l'ADPG montre une activité d'élongation pour la glgA, en haut du gel. Cette bande est également visible dans les extraits solubles contenant les protéines tronquées SS4.

Enfin, dans les gels d'activité contenant dans les extraits solubles de SS4∆349, 403 et 466, on observe deux bandes d'activité d'élongation (une d'intensité plus forte et une plus faible), autre que celle liée à glgA, Cela signifie que même dans sa forme tronquée, SS4 est capable d'allonger les chaînes glycanniques du glycogène à partir de l'ADPG, parce que le domaine catalytique est conservé. Il est également possible que ces deux bandes traduisent d'une multimérisation de SS4.

Comme toutes les SSs (sauf SS5), SS4 possède la capacité d'élongation, cependant elle a été décrite comme étant également impliquée dans les mécanismes d'initiation (Roldan et al.,2007). Cela signifie que SS4 serait capable d'initier l'élongation des chaînes de glucanes en utilisant seulement l'ADPG et des amorces de glucanes courtes disponibles dans le milieu. Dans l'optique de différencier l'activité d'élongation et l'activité d'initiation de SS4 nous avons réalisé un test d'activité, en utilisant le même protocole que précédemment mais en remplaçant le glycogène par du maltose ou un polymère de DP 7-13 dans la solution d'incubation avec l'ADPG. Cette expérience n'a pas révélé d'activité visible sur les gels (résultats non présentés). Ce qui signifie que dans ces conditions, SS4∆349, SS4∆403 et SS4∆466 n'ont pas réussi à initier l'élongation des chaînes. Le polymère DP-7-13 ressemble au glycogène en termes de longueur de chaîne et devrait donc être reconnue par la SS4. Nous ne pouvons donc pas savoir si l'absence d'activité provient du fait que le polyglucane n'arrive pas à rentrer dans le gel, que l'activité est trop faible pour être détectée sur ce type d'expérience ou si SS4, dans le gel n'arrive pas à capturer les sucres pour les allonger ou a besoin d'un autre partenaire de l'initiation (telles que PTST2 ou PII1).

5.2 PII1 inhiberait l'activité catalytique de SS4

Afin de vérifier si l'interaction de PII1 avec SS4 à un effet sur l'activité de SS4, nous avons mélangé les extraits solubles de PII1-H2H3 avec l'extrait soluble de SS4 Δ 349 et SS4 Δ 466. Nous avons choisi la construction SS4 Δ 349 la plus longue en considérant que les hélices plus nombreuses de SS4 puissent favoriser l'interaction avec PII1-H2H3. Nous avons également choisi la construction avec SS4 Δ 466 qui ne possède pas de CC mais seulement la région de dimérisation RD et le domaine catalytique afin de voir si la présence du domaine de dimérisation pourrait suffire pour l'interaction avec PII1-H2H3.

Pour cela, nous avons mélangé 175 μ g des extraits solubles bactériens exprimant PII1-H2H3 et SS4- Δ 349/ Δ 466 et incubé le mélange 30 min avant de le déposer sur le gel de polyacrylamide 10% avec 0,3% de glycogène. Comme précédemment, l'activité d'élongation de SS4 a été vérifiée par coloration à l'iode (**Figure 56**).

Pour ces tests, des contrôles rigoureux ont été mis en place. Comme précédemment, nous avons réalisé des expériences sans ADPG pour confirmer que l'activité observée était spécifique à ce substrat et déposé des lysats de bactéries BL21 non transformées pour vérifier que l'activité synthase provenait bien de SS4 et non pas des glgA ; L'expérience ici visant à démontrer un effet de PII1 sur l'activité de SS4, l'activité de glgA nous servira de témoin de la spécificité de cet effet.

Nous avons également déposé 175 µg de l'extrait soluble de PII1-H2H3 seul pour s'assurer de l'absence d'activité enzymatique chez cette protéine (Figure 59B).

Enfin, pour être certain que la baisse d'activité de SS4 n'est pas due à l'encombrement de PII1 présente en grande quantité et qui pourrait empêcher de fait l'action de la SS4, nous avons incubé SS4 avec la BSA en quantité très supérieurs à SS4 (20 mg/ml) (**Figure 56 B**).



Figure 56. Zymogramme révélant l'effet de PII1-H2-H3 sur l'activité d'élongation de SS4Δ349 et SS4Δ466. A) 175µg de l'extrait total de BL21(DE3) et SS4Δ349, SS4Δ466 ou PII1-H2H3 exprimée dans BL21(DE3) ont été charges sur gel de polyacrylamide non dénaturant 10% contenant 0.3% de glycogène d'huitre (Sigma). Après migration, les gels sont incubés toute la nuit à température ambiante dans un tampon d'incubation pour synthase avec 2,5 mM d'ADPG ou sans ADPG. L'activité des protéines sont révélées par incubation dans l'iode. B) Analyse zymogramme de l'effet de la BSA sur l'activité de SS4-Δ349.

L'analyse des résultats montre qu'en absence d'ADPG, aucune activité d'élongation n'est observée, ce qui est cohérent avec le fait que SS4 a besoin du précurseur comme substrat pour allonger les chaînes de glucanes.

Ensuite l'analyse des résultats des gels incubés avec de l'ADPG montre une activité d'élongation pour la glgA, en haut du gel. Cette bande est également visible dans les extraits solubles contenant les protéines tronquées SS4 et PII1-H2H3. Dans ces gels, PII1-H2H3 ne présente aucune activité d'élongation.

Enfin, dans les gels d'activité contenant dans les extraits solubles de SS4∆349 et SS4∆466 seule ou incubés avec PII1-H2H3, on observe deux bandes d'activité d'élongation (une d'intensité plus forte et une plus faible), autre que celle liée à glgA, En revanche en présence de PII1-H2H3, une forte inhibition de l'activité de SS4∆349 et SS4∆466 est observée. Cela signifie également que dans le cas du mélange avec SS4∆466, les régions CC ne sont pas requises pour l'interaction avec PII1-H2H3. Lors de cette expérience, j'ai également constaté que l'effet de PII1 nécessitait un temps d'incubation d'au minimum 15 min pour observer un effet sur l'activité de SS4.

Enfin la présence de la BSA à haute concentration n'a pas d'effet sur l'activité de SS4, Cela montre que l'effet inhibiteur de PII1 sur SS4 n'est pas dû à un simple encombrement, mais à une inhibition spécifique exercée par PII1.

Pour vérifier la spécificité de l'inhibition par PII1 nous avons reproduit l'expérience avec des quantités croissantes de PII1. Etant donné que nous travaillons à partir d'extraits solubles, nous étions en incapacité de déterminer les concentrations protéiques précises pour cela nous avons travaillé en termes de ratio protéine-protéine en se basant que les proportions relatives dans les extraits de PII1 et SS4 estimées sur SDS-PAGE (**Figure 57 A**). Pour cela nous avons graduellement augmenté la concentration de PII1 de 3 à 58 fois la quantité par rapport à SS4 (**Figure 56**).





Les résultats montrent une diminution progressive de l'intensité de l'activité synthase de SS4 à mesure de l'augmentation de la quantité de PII1-H2H3 (**Figure 57B**). Ces résultats confirment l'action dose dépendante de PII1 sur SS4. On estime que l'activité de SS4Δ349 diminue à partir de 12 fois la quantité de PII1-H2H3 par rapport à SS4Δ349 et plus la quantité de PII1 augmente, plus l'activité de SS4 diminue.

En résumé de ce chapitre, nous avons pu montrer pour la première fois l'activité d'élongation de SS4 bien que sa capacité à initier n'a pas pu être observé et montré un rôle complètement inattendu de PII1 en tant que répresseur spécifique de l'activité d'élongation de SS4.

DISCUSSION

Dans le cadre de la recherche de la fonction des protéines impliquées dans l'initiation visant à avancer dans la compréhension de l'initiation de la formation du grain d'amidon, ma thèse avait pour objectif de caractériser la fonction moléculaire de PII1, seule et en complexe avec SS4 son partenaire de l'initiation. Pour cela, nous avons développé plusieurs stratégies qui font appel à des approches pluridisciplinaires de biochimie, de biologie moléculaire, de biophysique et de bio-informatique.

La suppression de la région C-terminale favorise la solubilisation de PII1

Avant mon arrivée en thèse, des premiers tests d'expression ont été effectués sur PII1 et semblaient montrer que la protéine était exprimée et soluble. Dans un premier temps j'ai travaillé à partir de la construction disponible au laboratoire, qui exprimait PII1 entière sans le peptide de transit. Cette construction a rapidement montré son insolubilité sous la forme de corps d'inclusion. La présence des corps d'inclusion est plus fréquente chez les protéines eucaryotes exprimées chez *E.coli* par rapport à leur production *in vivo* et cela est principalement dû à leur niveau d'expression très élevé, rarement observé dans la nature (Wilkinson and Harrison., 1991). Il se pourrait aussi que PII1 ne puisse être stable qu'en présence des autres partenaires de l'initiation avec lesquels elle interagit, comme SS5,SS4 ou PTST2.

Toutes nos tentatives pour augmenter la solubilité de PII1 dans sa forme entière n'ont pas abouti, nous n'avons obtenu qu'une très faible quantité de protéines solubles, identifiée par spectrométrie de masse, mais insuffisante pour envisager en l'état une quelconque étude structurale par cristallographie ou SAXS.

Ces résultats nous ont conduit à réaliser des constructions tronquées en nous basant sur le modèle AF2. La structure était prédite entièrement en hélice α , plus exactement composée de 6 hélices de longueurs variables dont deux hélices H2 et H3 très longues (2/3 de la structure). Le score en pLDDT était très bon sur l'ensemble de la structure mais la matrice d'erreur semblait indiquer que les hélices étaient mal agencées les unes par rapport aux autres dans le modèle. Celui-ci prédisait des régions non repliées (situées en N- terminale) et entre les hélices H3 et H4. D'après ces observations nous avons réalisé quatre constructions tronquées. Cette stratégie a conduit à l'obtention de constructions protéiques exprimées et solubles et en particulier les protéines PII1-H2H3 et PII1-H1H2H3 montrant que la suppression de la région C-terminale favorise leur solubilisation.

PII1 adopte une structure dynamique organisée en coiled-coils

De la même façon qu'avec PII1 entière, nous avons optimisé l'expression, la solubilité et la purification de PII1-H2H3. De nombreuses conditions ont été testées. Le principal problème avec PII1-H2H3 n'était pas sa solubilité, mais plutôt l'hétérogénéité de l'échantillon protéique en solution. Nous avons remarqué que lors de la purification d'exclusion stérique sur une colonne de 25 ml, la protéine était éluée très tôt à 7 ml environ et une petite quantité de PII1 était également éluée tout au long du volume de la colonne. De plus lors de la concentration des protéines, une petite quantité de protéine (0,2-0,3 mg/ml) passait entre les filtres 0,25 de 30 kDa de seuil de poids moléculaire ce qui est en accord avec la structure AF3 de la protéine.

La structure prédite de PII1 par AF, montre une structure entièrement repliée en hélices avec une région désordonnée en N-terminal et entre H3 et H4. Les modèles de PII1- H2H3 que j'ai calculés avec AF3 et MassivFold par la suite montraient un repliement en CC des hélices H2 et H3. L'analyse par SR-CD du contenu en éléments de structure secondaire a confirmé la présence d'hélice agencées en CC. C'est probablement par la présence de ces CC que PII1 s'auto-assemble lors de l'expression en bactéries en l'absence de partenaires physiologiques.

Nous avons réalisé une expérience de SR-CD en présence d'une quantité croissante de TFE. Le TFE permet de créer un environnement hydrophobe qui déstabilise les CC pour favoriser la formation d'hélices simples. A faible concentration de TFE (10%) on a observé une déstructuration de la structure tertiaire de PII1-H2H3. Puis à partir de 30%, la protéine se replie progressivement sous la forme d'hélices simples pour se stabiliser à 50% de TFE. De façon étonnante, PII1 semble passer par plusieurs états selon le pourcentage de TFE ajouté. Cette expérience valide le fait que PII1 contient des CC mais également que PII1 possède une structure flexible et dynamique. Cela pourrait indiquer que PII1-H2H3 puisse adopter différents repliements en fonction du partenaire avec laquelle elle interagit, que ce soit SS4, SS5 ou PTST2 et à différents endroits de sa structure et transmettre des signaux via ces changements de conformation.

La région centrale de PII1 est conservée parmi les orthologues en séquence et en structure

Tous les orthologues PII1 extraits de la base de données TAIR présentent le même agencement en CC pour H2 et H3 (selon leur modèle AF). De plus, nous avons remarqué que l'orthologue *Setaria italica* possédait uniquement les hélices H1, H2 et H3, suggérant qu'elles sont nécessaires à la fonction de PII1. Nous avions mis en évidence des résidus très conservés (principalement des leucines) dans cette région montrant que ces hélices sont conservées. De plus, une comparaison de structure par AF3 a montré une similarité de structure entre les orthologues qui sont toutes prédites avec deux très longues hélices H2 et H3 associées en CC puis trois autres hélices plus petites en C-terminale dont l'ensemble forme une structure très allongée comme la myosine. La présence de ces CC renforce l'idée que les hélices de PII1 participent probablement à l'interaction avec des partenaires de l'initiation que ce soit avec SS4 (Vandromme et al., 2019), SS3 (Vandromme et al., 2023), PTST2 (Seung et al., 2018) ou SS5 (Sharma et al., 2024).

La région C-terminale de PII1 est également conservée

Grâce à l'alignement de séquence, nous avons également mis en évidence des résidus conservés (leucine et acide glutamique) dans la région C-terminale (H4, H5 et H6). Bien que nous n'ayons pas gardé cette région pour nos constructions tronquées, il est possible qu'elle joue un rôle que qu'il n'a pas été possible de caractériser avec nos données expérimentales.

On peut faire une analogie avec les travaux de Osman et al., 2024 dans lesquels j'ai participé. Le modèle AF3 de LESV prédisait un repliement d'une région composée d'hélices

a. Des expériences de SR-CD ont montré que certaines hélices apparaissent suite à la fixation de l'amylopectine sur la molécule. De la même façon que pour LESV, on peut émettre l'hypothèse que certaines hélices prédites par AF puissent ne se former que dans le cadre de la présence de partenaire et apporter ainsi une certaine plasticité à la molécule.

PII1 reconnait le site de dimérisation conservé de SS4 et SS5

SS4 et SS5 sont toutes les deux décrites comme étant des protéines dimériques en solution avec une région de dimérisation essentielle pour la dimérisation et l'activité, dans le cas de SS4 (Abt et al., 2020; Raynaud et al., 2016). Pour SS5, cette dimérisation est nécessaire à l'interaction avec PII1 (Abt et al., 2020). Nous avons montré par nos modèles AF, que la structure de la région de dimérisation RD, présente chez SS5 et SS4 est conservée dans les deux modèles. Elle est globulaire, composée de quatre hélices et interagit avec la sous unité GT5 du domaine catalytique. Dans les deux modèles SS5 et SS4 en dimère, la région de dimérisation du premier monomère interagit avec le GT5 du second monomère et inversement. Cet agencement permet de maintenir les deux domaines GT5 proche l'un de l'autre favorisant l'accessibilité de la région catalytique, dans le cas de SS4, ou du domaine de liaison aux glucanes dans le cas de SS5. Cette prédiction est validée par le score de pLDDT et le PAE.

En plus de la région RD, nous avons montré que la structure de SS5 se superpose avec le GT5 et avec l'hélice C-terminale rattachée au GT1 de SS4. Chez SS5 cette hélice C- terminale est rattachée au GT5 et reproduit ainsi le même environnement pour le GT5 que dans SS4. Avec nos analyses nous ne pouvons pas déterminer la fonction de cette hélice C-terminale mais il est probable que sa conservation structurelle chez SS4 et SS5 indique qu'elle peut jouer un rôle, peut-être de reconnaissance pour un partenaire.

Le point commun entre les deux protéines SS4 et SS5 est qu'elles interagissent toutes les deux avec PII1. PII1 n'interagit pas avec les autres SSs, et comme toutes les SSs possèdent un domaine catalytique (GT1 et GT5) on peut émettre l'hypothèse que ce n'est pas cette région qui est impliquée dans l'interaction avec PII1. Un autre point commun entre SS4 et SS5 (et la différence avec toutes les autres SSs) est qu'elles dimérisent via des domaines de dimérisation. Ceux-ci forment une région charnière conservée entre le domaine GT5 et la région CC. On peut logiquement envisager que c'est via cette région de dimérisation que SS4 et SS5 interagissent avec PII1. Ce point est confirmé par le fait que PII1 est capable d'interagir avec SS4 Δ 466 qui ne contient que le domaine catalytique et de dimérisation et d'inhiber son activité d'élongation.

Ces résultats, associés au fait que sans le domaine de dimérisation RD, SS5 n'est plus dimérique et ne reconnait pas PII1, conforte l'idée que PII1 interagit probablement avec SS4 et SS5 proche de cette région RD-GT5 (Abt et al., 2020). Ainsi il est également probable que PII1 interagit avec l'une ou l'autre (SS4 ou SS5) mais pas les deux simultanément. Cela pourrait entraîner une compétition entre SS4 et SS5 pour l'interaction avec PII1.

PII1 inhibe l'activité d'élongation de SS4

Dans ces travaux de thèse, l'activité d'élongation de SS4 a été visualisée sur zymogramme à partir du glycogène et de l'ADPG. Ces résultats montrent que la capacité de SS4 à allonger les sucres est fonctionnelle et que la partie catalytique avec le domaine de dimérisation sont suffisantes pour que SS4 soit active.

SS4 est l'enzyme principale de l'initiation, mais sa fonction d'initiation n'a jamais été observée sur gel d'activité. Dans mon travail de thèse j'ai essayé de reproduire l'étape d'initiation en déposant les constructions tronquées sur un gel polyacrylamide sans glycogène et en ajoutant les sucres (maltoses et polyglucanes DP 7). Ces sucres miment les MOS libres dans les conditions physiologiques de la plante, utilisés par SS4 lors de l'initiation. Dans la plante, il a été montré que SS4 peut reconnaitre le maltose, mais a une préférence pour les MOS plus longs (Pfister and Zeeman., 2016). En rajoutant les maltooligosides de DP 7-13, nous avons mimé les chaînes courtes de glycogène composées d'au moins sept glucoses. Dans ces conditions, nous n'avons pas observé d'activité d'élongation et, donc, pas d'initiation. Nous pouvons expliquer ce résultat de plusieurs façons : Premièrement il est possible que l'initiation ait eu lieu mais que l'activité soit trop faible pour être détectée. Les gels d'activité ont été révélés à l'iode ou avec le réactif de Schiff qui permet de détecter les sucres de façon plus sensible. Il est également possible

que les sucres ajoutés n'arrivent pas à se fixer dans le gel. Deuxièmement, le mécanisme d'initiation est complexe et fait appel à au moins six protéines : SS4, PTST2, PTST3, SS5, MFP1 et PII1. Il est possible qu'il faille plus que SS4 pour que son activité d'initiation soit optimale. Troisièmement, étant donné que nous travaillons sur des constructions tronquées et en condition non physiologique, il est probable que SS4 ait besoin de ces CC pour avoir une activité d'initiation optimale.

Comment expliquer le phénotype du mutant pii1 chez Arabidopsis thaliana?

La suppression de PII1 dans la plante, conduit à un phénotype avec un grain par chloroplaste. Ces grains sont plus gros que dans la plante sauvage mais conservent leur forme lenticulaire contrairement aux mutants *ss4* (Seung et al., 2018). A première vue, les hypothèses à l'égard de PII1 portaient principalement sur une régulation des évènements d'initiation mais peu d'autres expériences, mis à part la génétique inverse et les interactions protéiques, n'ont permis d'avancer dans la compréhension de son rôle au sein du complexe d'initiation (Vandromme et al., 2019, Seung et al., 2018, Vandromme et al., 2023). D'après mes recherches, j'ai montré qu'en solution, PII1 inhibe l'activité d'élongation de SS4. Dans ce cas on pourrait s'attendre à ce qu'en l'absence de PII1, le nombre de grain par chloroplaste soit plus important que chez la plante sauvage, alors comment expliquer le mutant *pii1* ?

Si on se replace dans le contexte de l'initiation, SS4 est capable d'initier la synthèse des amorces glycaniques qui seront ensuite prises en charge par les autres enzymes (SS, SBE et DBE) pour former un grain. En absence de SS4, la synthèse d'amidon est altérée car la plupart des chloroplastes sont sans amidon donc SS4 est indispensable à l'initiation (Crumpton-Taylor et al., 2013; Roldán et al., 2007; Szydlowski et al., 2009). Mon hypothèse est que SS4 allonge une amorce MOS d'un DP suffisamment grand puis PII1 inhibe l'élongation de SS4 à partir d'une certaine longueur de chaîne. Cette inhibition serait transitoire et n'interviendrait qu'à un moment précis dans le mécanisme de l'initiation afin de permettre l'allongement d'une autre amorce par SS4. Une fois inhibition de PII1

suspendu ou inactivée, SS4 recommencerait ensuite l'allongement d'une nouvelle amorce aboutissant à un nouveau grain.

En résumé en l'absence de PII1, SS4 continuerait à allonger son amorce sans arrêt pour ne former qu'un seul grain par chloroplaste, plus gros. PII1 pourrait en effet empêcher la l'activité de SS4 de façon transitoire, en bloquant son site actif localisé entre les deux domaines GT ou bien en empêchant sa dimérisation en se localisant proche de la région RD-GT5.

Proposition du mécanisme d'initiation

Les mécanismes de l'initiation de la biosynthèse de l'amidon sont complexes et certainement dynamiques. Ils impliquent de nombreuses interactions protéine-protéine mais on ne connait ni la taille du complexe, ni la stœchiométrie ou la stabilité. D'après l'état actuel des connaissances et nos travaux, il apparait clairement que certaines de ces interactions soient plutôt transitoires. On compte désormais six protéines impliquées dans l'initiation : SS4, SS5, PII1, PTST2, PTST3 et MFP1.

Le rôle de SS4 dans l'initiation est important, elle est considérée comme une enzyme dimérique majeure de l'initiation et capable d'allonger les chaînes de glucanes grâce à son domaine catalytique GT1 et GT5 (Roldán et al., 2007).

Il a été montré que SS4 interagit avec le CBM48 de PTST2. PTST2 se lie préférentiellement aux glucanes plus longs dont la structure est plutôt hélicoïdale que linéaire. Il a été supposé que, par le biais de son CBM48, PTST2 apporte les MOS à SS4, qui ne contient pas de CBM, qu'elle récupère et allonge ensuite (Seung et al., 2017). En revanche dans un modèle de levure exprimant différentes combinaisons d'enzymes impliquées dans la biosynthèse de l'amidon, la présence de SS4 favorisait la formation de glucanes insolubles, même en l'absence de PTST2 (Pfister et Zeeman, 2016). On peut se demander si l'étape d'initiation est réellement pertinente, si les autres SSs peuvent récupérer les MOS dans le stroma du chloroplaste. Or, PTST2 est indispensable pour initier la biosynthèse de l'amidon et est également responsable de la polyinitiation (augmentation

du nombre de petits grains par chloroplaste) lorsqu'elle est surexprimée chez les plantes. Une autre hypothèse peut être mis en avant quant à la fonction de PTST2. PTST2 pourrait non pas donner les substrats MOS à SS4 mais plutôt récupérer ses amorces hélicoïdales plus longues synthétisés par SS4 pour les rendre disponible pour les autres SSs.

Ensuite PII1 interviendrait dans initiation de la façon dont on l'a décrite précédemment c'est à dire en inhibant l'élongation des chaînes de glucanes par interaction avec SS4. A ce moment on peut supposer que PTST2 une fois le MOS libéré se repositionne pour supprimer l'inhibition par PII1 et permettre l'initiation d'une nouvelle amorce.

Le rôle de SS5 dans ce complexe d'initiation est encore flou. Le mutant *ss5* est beaucoup moins sévère que *ss4* car les chloroplastes comportent entre 1 et 3 grains (Abt et al., 2020). SS5 pourrait servir à stabiliser PII1 lorsqu'il n'est pas en interaction avec SS4 ou bien fixer les glucanes. D'après mes analyses, SS5 interagirait avec PII1 au niveau de son domaine de dimérisation/ GT5 et cela jusqu'au moment où PII1 doit interagir avec SS4 pour l'inhiber. En résume, PII1 interagit ou bien avec SS5 ou bien SS4 mais pas simultanément.

Enfin, MFP1 permet de guider l'ensemble du mécanisme primaire d'initiation proche des membranes des thylakoïdes en interagissant spécifiquement avec PTST2 (Sharma et al., 2024).

Concernant les protéines qui peuvent participer aux mécanismes d'initiation, j'inclurai également SS3, qui peut remplacer partiellement l'activité de SS4 en son absence (Szydlowski et al., 2009). J'ai montré que SS3 présente un domaine catalytique GT1 et GT5 qui lui permet d'allonger les sucres comme les autres synthases, mais elle présente également trois CMB53 reliés à une hélice. Cette structure ressemble à celle retrouvée chez PTST2 (et PTST3) qui possèdent un CBM48 et une hélice. Nous avons montré que PTST2 serait capable à l'aide de son CBM48, de récupérer les amorces allongées par SS4. Dans cette configuration, SS3 pourrait à la fois allonger cette amorce à l'aide de son site catalytique et mimer le rôle de PTST2 c'est-à-dire de délivrer les amorces aux autres enzymes de la biosynthèse de l'amidon (SSs, BEs et DBEs).

Le mécanisme d'initiation que je propose inclue les six protéines de l'initiation et est illustré en **Figure 58**. Ces mécanismes restent loin d'être éclaircis et il ne serait pas étonnant

167

que d'autres protéines encore indéterminées soient impliquées directement dans ce processus.



Figure 58. Proposition du mécanisme d'initiation de la biosynthèse du grain d'amidon chez la plante. 1) *Fixation du MOS sur SS4. 2) Allongement du MOS par l'activité d'élongation de SS4. 3) PII1 se place sur SS4 de façon à changer sa conformation et arrêter l'allongement des MOS qui sera récupéré par PTST2 puis libéré dans le stroma du chloroplaste. 4) Le MOS libéré est pris en charge par les autres enzymes de la biosynthèse pour la formation d'un nouveau grain. PTST2 se repositionne de façon à ce que SS4 puisse récupérer une nouvelle amorce. Nous ne connaissons pas le moment où intervient l'interaction entre SS5 et PII1 ou entre PTST2 et PTST3, ce qui est symbolisé par un point d'interrogation.*

PERSPECTIVES

Malgré les avancées réalisées dans l'étude du complexe protéique impliqué dans l'initiation des grains d'amidon, de nombreuses questions restent en suspens. Les connaissances actuelles se résument à l'étude des phénotypes des mutants surexprimant ou inactivant les protéines ainsi que l'identification des partenaires dans des expériences d'immunoprécipitation et double-hybride, mais ces analyses ne permettent pas d'aller plus loin dans l'interprétation. Actuellement, il n'est pas encore clairement établi si les protéines de l'initiation interagissent simultanément ou de manière séquentielle au cours de l'étape d'initiation ou quelle est la taille et la stœchiométrie du complexe.

Dans mon travail, j'ai étudié les aspects structurels et fonctionnel de PII1 et SS4.

J'ai pu montrer une inhibition de l'activité de SS4 par PII1. Ce résultat extrêmement nouveau a permis de progresser dans la compréhension du mécanisme d'initiation et d'émettre de nouvelles hypothèses. Cependant, ces expériences ont été faites sur des constructions tronquées et surtout en l'absence des autres partenaires. Pour progresser dans nos hypothèses et notamment sur l'articulation des rôles de PII1 et des autres partenaires, il serait intéressant de tester l'effet de la présence de PTST2, SS5 et PII1 sur l'activité de SS4 ainsi que la compétition éventuelle entre les différentes protéines de l'initiation. Pour cela le zymogramme parait particulièrement bien adapté pour des protéines difficiles à produire et permettrait à court terme d'obtenir des informations supplémentaires.

La surexpression de PTST2 chez *Arabidopsis* a montré une sur-nucléation des grains d'amidon attestant d'une sur-initiation. La surexpression de PII1 n'a pas été testée et pourrait donner des informations sur le mécanisme de PII1 et confirmer l'antagonisme de PII1 et PTST2.

L'approche structurale par AlphaFold a donné une dimension supplémentaire à l'étude de l'initiation. Ainsi nous avons montré que les domaines de dimérisation des SS4 et SS5 ne présentent pas d'identité de séquence mais une forte identité de structure. Cela nous a permis de proposer un mode de dimérisation commun aux deux protéines et de localiser une région d'interaction commune avec PII1.

L'analyse des modèles moléculaires de tous les partenaires de l'initiation a permis de visualiser un grand nombre de régions CC ou désordonnées attestant de l'aspect dynamique des différents mécanismes moléculaires impliqués et augmentant le nombre de possibilités d'association, de changements de conformation et de diffusion des signaux associés.

Pour le moment, AlphaFold n'a pas permis de modéliser les interactions entre les partenaires mais les progrès galopants des approches par AI devraient permettre de progresser rapidement.

D'un point de vue structural, le complexe PII1/SS4 semble instable et hétérogène, ce qui rend difficile son étude par cristallographie aux rayons X, notamment en raison de la formation d'agrégats solubles. Il est probable que l'ajout d'un troisième partenaire (PTST2), voire davantage (SS5), serait nécessaire pour stabiliser ce complexe. La prochaîne étape serait donc d'explorer les interactions de PII1 et SS4 avec PTST2 et SS5 afin de tester la stabilité du complexe et d'envisager des études structurales plus approfondies, telles que la cristallographie aux rayons X ou le SAXS voire la cryo-microscopie électronique (cryo-EM). MFP1, en raison de sa structure allongée et difficile à produire en bactérie et de son rôle établi dans la localisation du complexe d'initiation, pourrait être écartée de ces analyses structurales.

Il est clair que la complexité des mécanismes de l'initiation est importante faisant appel à de nombreux échanges moléculaires. Il est possible voire vraisemblable que certains partenaires transitoires n'aient pas encore pu être identifiés par les approches utilisées, notamment des partenaires qui feraient le lien avec la machinerie enzymatique permettant l'élongation des chaînes d'amylopectine.

171

Bilan fin de thèse

Ce projet de thèse s'est révélé particulièrement exigeant. En effet, la production de la protéine PII1, par exemple, a présenté des défis inattendus de par son insolubilité. Bien que la production de constructions tronquées n'ait pas été initialement prévue dans le cadre de mes travaux, cette difficulté s'est avérée formatrice. Elle m'a permis d'acquérir des compétences en clonage pour réaliser ces constructions, ainsi qu'en bioinformatique pour modéliser les structures. La complexité du projet m'a également contraint à approfondir ma réflexion scientifique et de gagner en autonomie dans l'expression des protéines recombinantes. De plus, j'ai eu l'opportunité de finaliser le projet de Rayan Osman, dont l'objectif était de caractériser la fonction des protéines non enzymatiques ESV1 et LESV. L'ensemble de ces projets a considérablement enrichi ma formation doctorale d'un point de vue technique et scientifique.

REFERENCES

- Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., Bodenstein, S.W., Evans, D.A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A.I., Cowie, A., Figurnov, M., Fuchs, F.B., Gladman, H., Jain, R., Khan, Y.A., Low, C.M.R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E.D., Zielinski, M., Žídek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D., Jumper, J.M., 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630, 493–500. https://doi.org/10.1038/s41586-024-07487-w
- Abt, M.R., Pfister, B., Sharma, M., Eicke, S., Bürgy, L., Neale, I., Seung, D., Zeeman, S.C., 2020. STARCH SYNTHASE5, a Noncanonical Starch Synthase-Like Protein, Promotes Starch Granule Initiation in Arabidopsis. Plant Cell 32, 2543–2565. https://doi.org/10.1105/tpc.19.00946
- Abt, M.R., Zeeman, S.C., 2020. Evolutionary innovations in starch metabolism. Current Opinion in Plant Biology 55, 109–117. https://doi.org/10.1016/j.pbi.2020.03.001
- Ahmed, Z., Tetlow, I.J., Ahmed, R., Morell, M.K., Emes, M.J., 2015. Protein–protein interactions among enzymes of starch biosynthesis in high-amylose barley genotypes reveal differential roles of heteromeric enzyme complexes in the synthesis of A and B granules. Plant Science 233, 95–106. https://doi.org/10.1016/j.plantsci.2014.12.016
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of Molecular Biology 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
- Apriyanto, A., Compart, J., Fettke, J., 2022. A review of starch, a unique biopolymer Structure, metabolism and in planta modifications. Plant Science 318, 111223. https://doi.org/10.1016/j.plantsci.2022.111223
- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res 44, W344–W350. https://doi.org/10.1093/nar/gkw408
- Bahaji, A., Li, J., Ovecka, M., Ezquer, I., Muñoz, F.J., Baroja-Fernández, E., Romero, J.M., Almagro, G., Montero, M., Hidalgo, M., Sesma, M.T., Pozueta-Romero, J., 2011. Arabidopsis thaliana Mutants Lacking ADP-Glucose Pyrophosphorylase Accumulate Starch and Wild-type ADP-Glucose Content: Further Evidence for the Occurrence of Important Sources, other than ADP-Glucose Pyrophosphorylase, of ADP-Glucose Linked to Leaf Starch Biosynthesis. Plant and Cell Physiology 52, 1162–1176. https://doi.org/10.1093/pcp/pcr067
- Ball, S., Guan, H.-P., James, M., Myers, A., Keeling, P., Mouille, G., Buléon, A., Colonna, P., Preiss, J., 1996. From Glycogen to Amylopectin: A Model for the Biogenesis of the Plant Starch Granule. Cell 86, 349–352. https://doi.org/10.1016/S0092-8674(00)80107-5
- Ballicora, M.A., Frueauf, J.B., Fu, Y., Schürmann, P., Preiss, J., 2000. Activation of the Potato Tuber ADP-glucose Pyrophosphorylase by Thioredoxin. Journal of Biological Chemistry 275, 1315–1320. https://doi.org/10.1074/jbc.275.2.1315
- Banks, William, Banks, W., Greenwood, C. T., Greenwood, Charles T., 1975. Starch and its components. Edinburgh Univ. Pr, Edinburgh.

- Barengo, R., Flawiá, M., Krisman, C.R., 1975. The initiation of glycogen biosynthesis in *Escherichia coli*. FEBS Letters 53, 274–278. https://doi.org/10.1016/0014-5793(75)80035-4
- Bertoft, E., 1991. Chains of intermediate lengths in waxy-maize amylopectin. Carbohydrate Research 212, 245–251. https://doi.org/10.1016/0008-6215(91)84061-l
- Biliaderis, G., 1979. A Comparison of the Enzymatic Hydrolysis of Smooth Pea Starch to that of Corn and Wheat. Canadian Institute of Food Science and Technology Journal 12, 131–134. https://doi.org/10.1016/S0315-5463(79)73099-9
- Boyer, C.D., Preiss, J., 1978. Multiple forms of $(1 \rightarrow 4)$ - α -d-glucan, $(1 \rightarrow 4)$ - α -d-glucan-6-glycosyl transferase from developing zea mays L. Kernels. Carbohydrate Research 61, 321–334. https://doi.org/10.1016/S0008-6215(00)84492-4
- Brust, H., Lehmann, T., D'Hulst, C., Fettke, J., 2014. Analysis of the Functional Interaction of Arabidopsis Starch Synthase and Branching Enzyme Isoforms Reveals that the Cooperative Action of SSI and BEs Results in Glucans with Polymodal Chain Length Distribution Similar to Amylopectin. PLoS ONE 9, e102364. https://doi.org/10.1371/journal.pone.0102364
- Brysbaert, G., Raouraoua, N., Mirabello, C., Véry, T., Blanchet, C., Wallner, B., Lensink, M., 2024. MassiveFold: unveiling AlphaFold's hidden potential with optimized and parallelized massive sampling. https://doi.org/10.21203/rs.3.rs-4319486/v1
- Buléon, A., Colonna, P., Planchot, V., Ball, S., 1998. Starch granules: structure and biosynthesis. International Journal of Biological Macromolecules 23, 85–112. https://doi.org/10.1016/S0141-8130(98)00040-3
- Bürgy, L., Eicke, S., Kopp, C., Jenny, C., Lu, K.J., Escrig, S., Meibom, A., Zeeman, S.C., 2021. Coalescence and directed anisotropic growth of starch granule initials in subdomains of Arabidopsis thaliana chloroplasts. Nat Commun 12, 6944. https://doi.org/10.1038/s41467-021-27151-5
- Burkhard, P., Stetefeld, J., Strelkov, S.V., 2001. Coiled coils: a highly versatile protein folding motif. Trends in Cell Biology 11, 82–88. https://doi.org/10.1016/S0962-8924(00)01898-5
- Calvin, M., Benson, A.A., 1948. The Path of Carbon in Photosynthesis. Science 107, 476–480. https://doi.org/10.1126/science.107.2784.476
- Crofts, N., Abe, N., Oitome, N.F., Matsushima, R., Hayashi, M., Tetlow, I.J., Emes, M.J., Nakamura, Y., Fujita, N., 2015. Amylopectin biosynthetic enzymes from developing rice seed form enzymatically active protein complexes. EXBOTJ 66, 4469–4482. https://doi.org/10.1093/jxb/erv212
- Crumpton-Taylor, M., Grandison, S., Png, K.M.Y., Bushby, A.J., Smith, A.M., 2012. Control of Starch Granule Numbers in Arabidopsis Chloroplasts. Plant Physiology 158, 905– 916. https://doi.org/10.1104/pp.111.186957
- Crumpton-Taylor, M., Pike, M., Lu, K., Hylton, C.M., Feil, R., Eicke, S., Lunn, J.E., Zeeman, S.C., Smith, A.M., 2013. Starch synthase 4 is essential for coordination of starch granule formation with chloroplast division during Arabidopsis leaf expansion. New Phytol 200, 1064–1075. https://doi.org/10.1111/nph.12455
- Cuesta-Seijo, J.A., Nielsen, M.M., Ruzanski, C., Krucewicz, K., Beeren, S.R., Rydhal, M.G., Yoshimura, Y., Striebeck, A., Motawia, M.S., Willats, W.G.T., Palcic, M.M., 2016. In vitro Biochemical Characterization of All Barley Endosperm Starch Synthases. Front. Plant Sci. 6. https://doi.org/10.3389/fpls.2015.01265

Czaja, A.T., 1969. Die Mikroskopie der Stärkekörner. In: Handbuch der Stärke in

Einzeldarstellungen. Berlin

- Delatte, T., Trevisan, M., Parker, M.L., Zeeman, S.C., 2005. Arabidopsis mutants Atisa1 and Atisa2 have identical phenotypes and lack the same multimeric isoamylase, which influences the branch point distribution of amylopectin during starch synthesis: Isoamylase function in Arabidopsis. The Plant Journal 41, 815–830. https://doi.org/10.1111/j.1365-313X.2005.02348.x
- Delorenzi, M., Speed, T., 2002. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. Bioinformatics 18, 617–625. https://doi.org/10.1093/bioinformatics/18.4.617
- Delvallé, D., Dumez, S., Wattebled, F., Roldán, I., Planchot, V., Berbezy, P., Colonna, P., Vyas,
 D., Chatterjee, M., Ball, S., Mérida, Á., D'Hulst, C., 2005. Soluble starch synthase I: a major determinant for the synthesis of amylopectin in Arabidopsis thaliana leaves: SSI mutant of Arabidopsis. The Plant Journal 43, 398–412. https://doi.org/10.1111/j.1365-313X.2005.02462.x
- Denyer, K., Waite, D., Motawia, S., Møller, B.L., Smith, A.M., 1999. Granule-bound starch synthase I in isolated starch granules elongates malto-oligosaccharides processively. Biochemical Journal 340, 183–191. https://doi.org/10.1042/bj3400183
- Dumez, S., Wattebled, F., Dauvillee, D., Delvalle, D., Planchot, V., Ball, S.G., D'Hulst, C., 2006. Mutants of *Arabidopsis* Lacking Starch Branching Enzyme II Substitute Plastidial Starch Synthesis by Cytoplasmic Maltose Accumulation. The Plant Cell 18, 2694–2709. https://doi.org/10.1105/tpc.105.037671
- Echt, C.S., Schwartz, D., 1981. EVIDENCE FOR THE INCLUSION OF CONTROLLING ELEMENTS WITHIN THE STRUCTURAL GENE AT THE WAXY LOCUS IN MAIZE. Genetics 99, 275–284. https://doi.org/10.1093/genetics/99.2.275
- Feike, D., Seung, D., Graf, A., Bischof, S., Ellick, T., Coiro, M., Soyk, S., Eicke, S., Mettler-Altmann, T., Lu, K.J., Trick, M., Zeeman, S.C., Smith, A.M., 2016. The Starch Granule-Associated Protein EARLY STARVATION1 Is Required for the Control of Starch Degradation in *Arabidopsis thaliana* Leaves. Plant Cell 28, 1472–1489. https://doi.org/10.1105/tpc.16.00011
- Fisher, D.K., Boyer, C.D., Hannah, L.C., 1993. Starch Branching Enzyme II from Maize Endosperm. Plant Physiol. 102, 1045–1046. https://doi.org/10.1104/pp.102.3.1045
- Fulton, D.C., Stettler, M., Mettler, T., Vaughan, C.K., Li, J., Francisco, P., Gil, M., Reinhold, H., Eicke, S., Messerli, G., Dorken, G., Halliday, K., Smith, A.M., Smith, S.M., Zeeman, S.C., 2008. β-AMYLASE4, a Noncatalytic Protein Required for Starch Breakdown, Acts Upstream of Three Active β-Amylases in *Arabidopsis* Chloroplasts. The Plant Cell 20, 1040–1058. https://doi.org/10.1105/tpc.107.056507
- Furukawa, K., Tagaya, M., Inouye, M., Preiss, J., Fukui, T., 1990. Identification of lysine 15 at the active site in Escherichia coli glycogen synthase. Conservation of Lys-X-Gly-Gly sequence in the bacterial and mammalian enzymes. J Biol Chem 265, 2086–2090.
- Gámez-Arjona, F.M., Raynaud, S., Ragel, P., Mérida, Á., 2014. Starch synthase 4 is located in the thylakoid membrane and interacts with plastoglobule-associated proteins in Arabidopsis. Plant J 80, 305–316. https://doi.org/10.1111/tpj.12633
- Gao, M., Fisher, D.K., Kim, K.N., Shannon, J.C., Guiltinan, M.J., 1997. Independent Genetic Control of Maize Starch-Branching Enzymes IIa and IIb (Isolation and

Characterization of a Sbe2a cDNA). Plant Physiology 114, 69–78. https://doi.org/10.1104/pp.114.1.69

- Garnier, J., Gibrat, J.-F., Robson, B., 1996. [32] GOR method for predicting protein secondary structure from amino acid sequence, in: Methods in Enzymology. Elsevier, pp. 540–553. https://doi.org/10.1016/S0076-6879(96)66034-0
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein Identification and Analysis Tools on the ExPASy Server, in: Walker, J.M. (Ed.), The Proteomics Protocols Handbook. Humana Press, Totowa, NJ, pp. 571–607. https://doi.org/10.1385/1-59259-890-0:571
- Gibon, Y., Bläsing, O.E., Palacios-Rojas, N., Pankovic, D., Hendriks, J.H.M., Fisahn, J., Höhne, M., Günther, M., Stitt, M., 2004. Adjustment of diurnal starch turnover to short days: depletion of sugar during the night leads to a temporary inhibition of carbohydrate utilization, accumulation of sugars and post-translational activation of ADP-glucose pyrophosphorylase in the following light period. The Plant Journal 39, 847–862. https://doi.org/10.1111/j.1365-313X.2004.02173.x
- Grimaud, F., Rogniaux, H., James, M.G., Myers, A.M., Planchot, V., 2008. Proteome and phosphoproteome analysis of starch granule-associated proteins from normal maize and mutants affected in starch biosynthesis. Journal of Experimental Botany 59, 3395–3406. https://doi.org/10.1093/jxb/ern198
- Gruber, M., Söding, J., Lupas, A.N., 2006. Comparative analysis of coiled-coil prediction methods. Journal of Structural Biology 155, 140–145. https://doi.org/10.1016/j.jsb.2006.03.009
- Harn, C., Knight, M., Ramakrishnan, A., Guan, H., Keeling, P.L., Wasserman, B.P., 1998. Isolation and characterization of the zSSIIa and zSSIIb starch synthase cDNA clones from maize endosperm. Plant Molecular Biology 37, 639–649. https://doi.org/10.1023/A:1006079009072
- Helle, S., Bray, F., Verbeke, J., Devassine, S., Courseaux, A., Facon, M., Tokarski, C., Rolando, C., Szydlowski, N., 2018. Proteome Analysis of Potato Starch Reveals the Presence of New Starch Metabolic Proteins as Well as Multiple Protease Inhibitors. Front. Plant Sci. 9, 746. https://doi.org/10.3389/fpls.2018.00746
- Hendriks, J.H.M., Kolbe, A., Gibon, Y., Stitt, M., Geigenberger, P., 2003. ADP-Glucose
 Pyrophosphorylase Is Activated by Posttranslational Redox-Modification in
 Response to Light and to Sugars in Leaves of Arabidopsis and Other Plant Species.
 Plant Physiology 133, 838–849. https://doi.org/10.1104/pp.103.024513
- Hizukuri, S., 1986. Polymodal distribution of the chain lengths of amylopectins, and its significance. Carbohydrate Research 147, 342–347. https://doi.org/10.1016/S0008-6215(00)90643-8
- Hizukuri, S., Takeda, Y., Yasuda, M., Suzuki, A., 1981. Multi-branched nature of amylose and the action of debranching enzymes. Carbohydrate Research 94, 205–213. https://doi.org/10.1016/S0008-6215(00)80718-1
- Hussain, H., Mant, A., Seale, R., Zeeman, S., Hinchliffe, E., Edwards, A., Hylton, C., Bornemann, S., Smith, A.M., Martin, C., Bustos, R., 2003. Three Isoforms of Isoamylase Contribute Different Catalytic Properties for the Debranching of Potato Glucans[W]. The Plant Cell 15, 133–149. https://doi.org/10.1105/tpc.006635
- James, M.G., Robertson, D.S., Myers, A.M., 1995. Characterization of the maize gene sugary1, a determinant of starch composition in kernels. Plant Cell 7, 417–429. https://doi.org/10.1105/tpc.7.4.417

- Jane, J.-L., Kasemsuwan, T., Leas, S., Zobel, H., Robyt, J.F., 1994. Anthology of Starch Granule Morphology by Scanning Electron Microscopy. Starch/Stärke 46, 121–129. https://doi.org/10.1002/star.19940460402
- Jenkins, P.J., Donald, A.M., 1995. The influence of amylose on starch granule structure. International Journal of Biological Macromolecules 17, 315–321. https://doi.org/10.1016/0141-8130(96)81838-1
- Jeong, S.Y., 2003. MFP1 is a thylakoid-associated, nucleoid-binding protein with a coiledcoil structure. Nucleic Acids Research 31, 5175–5185. https://doi.org/10.1093/nar/gkg693
- Jobling, S., 2004. Improving starch for food and industrial applications. Current Opinion in Plant Biology 7, 210–218. https://doi.org/10.1016/j.pbi.2003.12.001
- Jonasson, P., Liljeqvist, S., Nygren, P., Ståhl, S., 2002. Genetic design for facilitated production and recovery of recombinant proteins in *Escherichia coli*. Biotech and App Biochem 35, 91–105. https://doi.org/10.1042/BA20010099
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2
- Kaplan, F., Sung, D.Y., Guy, C.L., 2006. Roles of β-amylase and starch breakdown during temperatures stress. Physiologia Plantarum 126, 120–128. https://doi.org/10.1111/j.1399-3054.2006.00604.x
- Khalid, S., Yu, L., Meng, L., Liu, H., Ali, A., Chen, L., 2017. Poly(lactic acid)/starch composites: Effect of microstructure and morphology of starch granules on performance. Journal of Applied Polymer Science 134, 45504. https://doi.org/10.1002/app.45504
- Kirchberger, S., Leroch, M., Huynen, M.A., Wahl, M., Neuhaus, H.E., Tjaden, J., 2007. Molecular and Biochemical Analysis of the Plastidic ADP-glucose Transporter (ZmBT1) from Zea mays. Journal of Biological Chemistry 282, 22481–22491. https://doi.org/10.1074/jbc.M702484200
- Kofler, M.M., Freund, C., 2006. The GYF domain. The FEBS Journal 273, 245–256. https://doi.org/10.1111/j.1742-4658.2005.05078.x
- Kötting, O., Santelia, D., Edner, C., Eicke, S., Marthaler, T., Gentry, M.S., Comparot-Moss, S., Chen, J., Smith, A.M., Steup, M., Ritte, G., Zeeman, S.C., 2009. STARCH-EXCESS4 Is a Laforin-Like Phosphoglucan Phosphatase Required for Starch Degradation in *Arabidopsis thaliana*. The Plant Cell 21, 334–346. https://doi.org/10.1105/tpc.108.064360
- Leloir, L.F., Rongine de Fekete, M.A., Cardini, C.E., 1961. Starch and Oligosaccharide Synthesis from Uridine Diphosphate Glucose. Journal of Biological Chemistry 236, 636–641. https://doi.org/10.1016/S0021-9258(18)64280-2
- Liu, C., Pfister, B., Osman, R., Ritter, M., Heutinck, A., Sharma, M., Eicke, S., Fischer-Stettler, M., Seung, D., Bompard, C., Abt, M.R., Zeeman, S.C., 2023. LIKE EARLY STARVATION 1 and EARLY STARVATION 1 promote and stabilize amylopectin phase transition in starch biosynthesis. Sci. Adv. 9, eadg7448. https://doi.org/10.1126/sciadv.adg7448

- Liu, F., Makhmoudova, A., Lee, E.A., Wait, R., Emes, M.J., Tetlow, I.J., 2009. The amylose extender mutant of maize conditions novel protein-protein interactions between starch biosynthetic enzymes in amyloplasts. Journal of Experimental Botany 60, 4423–4440. https://doi.org/10.1093/jxb/erp297
- Lomako, J., Lomako, W.M., Whelan, W.J., 1988. A self-glucosylating protein is the primer for rabbit muscle glycogen biosynthesis. FASEB j. 2, 3097–3103. https://doi.org/10.1096/fasebj.2.15.2973423
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucl. Acids Res. 42, D490–D495. https://doi.org/10.1093/nar/gkt1178

Lu, K.-J., Pfister, B., Jenny, C., Eicke, S., Zeeman, S.C., 2018. Distinct Functions of STARCH SYNTHASE 4 Domains in Starch Granule Formation. Plant Physiol. 176, 566–581. https://doi.org/10.1104/pp.17.01008

Ludwiczak, J., Winski, A., Szczepaniak, K., Alva, V., Dunin-Horkawicz, S., 2019. DeepCoil-a fast and accurate prediction of coiled-coil domains in protein sequences. Bioinformatics 35, 2790–2795. https://doi.org/10.1093/bioinformatics/bty1062

Malinova, I., Alseekh, S., Feil, R., Fernie, A.R., Baumann, O., Schöttler, M.A., Lunn, J.E., Fettke, J., 2017. Starch Synthase 4 and Plastidal Phosphorylase Differentially Affect Starch Granule Number and Morphology. Plant Physiol. 174, 73–85. https://doi.org/10.1104/pp.16.01859

Malinova, I., Qasim, H.M., Brust, H., Fettke, J., 2018. Parameters of Starch Granule Genesis in Chloroplasts of Arabidopsis thaliana. Front. Plant Sci. 9, 761. https://doi.org/10.3389/fpls.2018.00761

Manners, D.J., 1989. Recent developments in our understanding of amylopectin structure. Carbohydrate Polymers 11, 87–112. https://doi.org/10.1016/0144-8617(89)90018-0

McGuffin, L.J., Bryson, K., Jones, D.T., 2000. The PSIPRED protein structure prediction server. Bioinformatics 16, 404–405. https://doi.org/10.1093/bioinformatics/16.4.404

Meyer (1883) Das Chlorophyllkorn in chemischer, morphologischer und biologischer Beziehung

Micsonai, A., Wien, F., Bulyáki, É., Kun, J., Moussong, É., Lee, Y.-H., Goto, Y., Réfrégiers, M., Kardos, J., 2018. BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. Nucleic Acids Research 46, W315–W322. https://doi.org/10.1093/nar/gky497

Miles, A.J., Wallace, B.A., 2018. CDtoolX, a downloadable software package for processing and analyses of circular dichroism spectroscopic data. Protein Science 27, 1717– 1722. https://doi.org/10.1002/pro.3474

Minelli, C., Liew, J.X., Muthu, M., Andresen, H., 2013. Coiled coil peptide-functionalized surfaces for reversible molecular binding. Soft Matter 9, 5119. https://doi.org/10.1039/c3sm50379h

Momma, M., Fujimoto, Z., 2012. Interdomain Disulfide Bridge in the Rice Granule Bound Starch Synthase I Catalytic Domain as Elucidated by X-Ray Structure Analysis. Bioscience, Biotechnology, and Biochemistry 76, 1591–1595. https://doi.org/10.1271/bbb.120305

Monroe, J.D., Storm, A.R., 2018. Review: The Arabidopsis β-amylase (BAM) gene family: Diversity of form and function. Plant Science 276, 163–170. https://doi.org/10.1016/j.plantsci.2018.08.016 Morrison, W.R., Milligan, T.P., Azudin, M.N., 1984. A relationship between the amylose and lipid contents of starches from diploid cereals. Journal of Cereal Science 2, 257–271. https://doi.org/10.1016/S0733-5210(84)80014-4

Mouille, G., Maddelein, M.L., Libessart, N., Talaga, P., Decq, A., Delrue, B., Ball, S., 1996. Preamylopectin Processing: A Mandatory Step for Starch Biosynthesis in Plants. Plant Cell 1353–1366. https://doi.org/10.1105/tpc.8.8.1353

Müller-Thurgau H (1882) Ueber zuckeranhäufung in pflanzentheilen in folge niederer temperatur.Landw Jahrb Suisse 11 : 751–828

- Myers, A.M., Morell, M.K., James, M.G., Ball, S.G., 2000. Recent Progress toward Understanding Biosynthesis of the Amylopectin Crystal1. Plant Physiology 122, 989– 998. https://doi.org/10.1104/pp.122.4.989
- Nakamura, Y., Umemoto, T., Ogata, N., Kuboki, Y., Yano, M., Sasaki, T., 1996. Starch debranching enzyme (R-enzyme or pullulanase) from developing rice endosperm: purification, cDNA and chromosomal localization of the gene. Planta 199. https://doi.org/10.1007/BF00196561
- Nelson, J.W., Kallenbach, N.R., 1986. Stabilization of the ribonuclease S-peptide a-helix by trifluoroethanol. Proteins 1, 211–217. https://doi.org/10.1002/prot.340010303
- Nelson, O.E., Tsai, C.Y., 1964. Glucose Transfer from Adenosine Diphosphate-Glucose to Starch in Preparations of Waxy Seeds. Science 145, 1194–1195. https://doi.org/10.1126/science.145.3637.1194
- Nielsen, M.M., Ruzanski, C., Krucewicz, K., Striebeck, A., Cenci, U., Ball, S.G., Palcic, M.M., Cuesta-Seijo, J.A., 2018. Crystal Structures of the Catalytic Domain of Arabidopsis thaliana Starch Synthase IV, of Granule Bound Starch Synthase From CLg1 and of Granule Bound Starch Synthase I of Cyanophora paradoxa Illustrate Substrate Recognition in Starch Synthases. Front. Plant Sci. 9, 1138. https://doi.org/10.3389/fpls.2018.01138
- Osman, R., Bossu, M., Dauvillée, D., Spriet, C., Liu, C., Zeeman, S.C., D'Hulst, C., Bompard, C., 2024. LIKE EARLY STARVATION 1 interacts with amylopectin during starch biosynthesis. Plant Physiology kiae193. https://doi.org/10.1093/plphys/kiae193
- O'Sullivan, A.C., Perez, S., 1999. The relationship between internal chain length of amylopectin and crystallinity in starch. Biopolymers 50, 381–390. https://doi.org/10.1002/(SICI)1097-0282(19991005)50:4<381::AID-BIP4>3.0.CO;2-W
- Pan, D., Nelson, O.E., 1984. A Debranching Enzyme Deficiency in Endosperms of the *Sugary-1* Mutants of Maize. Plant Physiol. 74, 324–328. https://doi.org/10.1104/pp.74.2.324
- Papadopoulos, J.S., Agarwala, R., 2007. COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics 23, 1073–1079. https://doi.org/10.1093/bioinformatics/btm076
- Peng, C., Wang, Y., Liu, F., Ren, Y., Zhou, K., Lv, J., Zheng, M., Zhao, S., Zhang, L., Wang, C., Jiang, L., Zhang, X., Guo, X., Bao, Y., Wan, J., 2014. *FLOURY ENDOSPERM6* encodes a CBM48 domain-containing protein involved in compound granule formation and starch synthesis in rice endosperm. Plant J 77, 917–930. https://doi.org/10.1111/tpj.12444
- Pfister, B., Zeeman, S.C., 2016. Formation of starch in plant cells. Cell. Mol. Life Sci. 73, 2781–2807. https://doi.org/10.1007/s00018-016-2250-x
- Qasba, P.K., Ramakrishnan, B., Boeggeman, E., 2005. Substrate-induced conformational changes in glycosyltransferases. Trends in Biochemical Sciences 30, 53–62. https://doi.org/10.1016/j.tibs.2004.11.005
- Raynaud, S., Ragel, P., Rojas, T., Mérida, Á., 2016. The N-terminal Part of Arabidopsis thaliana Starch Synthase 4 Determines the Localization and Activity of the Enzyme. Journal of Biological Chemistry 291, 10759–10771. https://doi.org/10.1074/jbc.M115.698332

Recondo, Leloir, 1961. ADENOSINE DIPHOSPHATE GLUCOSE AND STARCH SYNTHESIS.

- Ritte, G., Heydenreich, M., Mahlow, S., Haebel, S., Kötting, O., Steup, M., 2006. Phosphorylation of C6- and C3-positions of glucosyl residues in starch is catalysed by distinct dikinases. FEBS Letters 580, 4872–4876. https://doi.org/10.1016/j.febslet.2006.07.085
- Robert, X., Gouet, P., 2014. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Research 42, W320–W324. https://doi.org/10.1093/nar/gku316
- Roldán, I., Wattebled, F., Mercedes Lucas, M., Delvallé, D., Planchot, V., Jiménez, S., Pérez, R., Ball, S., D'Hulst, C., Mérida, Á., 2007. The phenotype of soluble starch synthase IV defective mutants of Arabidopsis thaliana suggests a novel function of elongation enzymes in the control of starch granule formation: Function of starch synthase type IV. The Plant Journal 49, 492–504. https://doi.org/10.1111/j.1365-313X.2006.02968.x
- Rongine De Fekete, M.A., Leloir, L.F., Cardini, C.E., 1960. Mechanism of Starch Biosynthesis. Nature 187, 918–919. https://doi.org/10.1038/187918a0
- Saito, M., Tanaka, T., Sato, K., Vrinten, P., Nakamura, T., 2018. A single nucleotide polymorphism in the "Fra" gene results in fractured starch granules in barley. Theor Appl Genet 131, 353–364. https://doi.org/10.1007/s00122-017-3006-1
- Santelia, D., Zeeman, S.C., 2011. Progress in Arabidopsis starch research and potential biotechnological applications. Current Opinion in Biotechnology 22, 271–280. https://doi.org/10.1016/j.copbio.2010.11.014
- Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D., Lohmann, J.U., 2005. A gene expression map of Arabidopsis thaliana development. Nat Genet 37, 501–506. https://doi.org/10.1038/ng1543
- Schrödinger, L., DeLano, W. 2020. PyMOL.
- Seung, D., Boudet, J., Monroe, J., Schreier, T.B., David, L.C., Abt, M., Lu, K.-J., Zanella, M.,

Zeeman, S.C., 2017. Homologs of PROTEIN TARGETING TO STARCH Control Starch

Granule Initiation in Arabidopsis Leaves. Plant Cell 29, 1657–1677.

- https://doi.org/10.1105/tpc.17.00222
- Seung, D., M. Smith, A., 2018. Starch granule initiation and morphogenesis- progress in Arabidopsis and cereals. https://doi.org/10.1093
- Seung, D., Schreier, T.B., Bürgy, L., Eicke, S., Zeeman, S.C., 2018. Two Plastidial Coiled-Coil Proteins Are Essential for Normal Starch Granule Initiation in Arabidopsis. Plant Cell 30, 1523–1542. https://doi.org/10.1105/tpc.18.00219
- Seung, D., Soyk, S., Coiro, M., Maier, B.A., Eicke, S., Zeeman, S.C., 2015. PROTEIN TARGETING TO STARCH Is Required for Localising GRANULE-BOUND STARCH

SYNTHASE to Starch Granules and for Normal Amylose Synthesis in Arabidopsis. PLoS Biol 13, e1002080. https://doi.org/10.1371/journal.pbio.1002080

- Sharma, M., Abt, M.R., Eicke, S., Ilse, T.E., Liu, C., Lucas, M.S., Pfister, B., Zeeman, S.C.,
 2024. MFP1 defines the subchloroplast location of starch granule initiation. Proc.
 Natl. Acad. Sci. U.S.A. 121, e2309666121.
 https://doi.org/10.1073/pnas.2309666121
- Smith, A.M., 2008. Prospects for increasing starch and sucrose yields for bioethanol production. The Plant Journal 54, 546–558. https://doi.org/10.1111/j.1365-313X.2008.03468.x
- Sprague, G.F., Brimhall, B., Hixon, R.M., 1943. Some Effects of the Waxy Gene in Corn on Properties of the Endosperm Starch ¹. Agronomy Journal 35, 817–822. https://doi.org/10.2134/agronj1943.00021962003500090008x
- Streb, S., Delatte, T., Umhang, M., Eicke, S., Schorderet, M., Reinhardt, D., Zeeman, S.C., 2009. Starch Granule Biosynthesis in *Arabidopsis* Is Abolished by Removal of All Debranching Enzymes but Restored by the Subsequent Removal of an Endoamylase. The Plant Cell 20, 3448–3466. https://doi.org/10.1105/tpc.108.063487
- Szydlowski, N., Ragel, P., Raynaud, S., Lucas, M.M., Roldán, I., Montero, M., Muñoz, F.J., Ovecka, M., Bahaji, A., Planchot, V., Pozueta-Romero, J., D'Hulst, C., Mérida, Á., 2009. Starch Granule Initiation in *Arabidopsis* Requires the Presence of Either Class IV or Class III Starch Synthases. The Plant Cell 21, 2443–2457. https://doi.org/10.1105/tpc.109.066522
- Tetlow, I.J., 2004. Recent developments in understanding the regulation of starch metabolism in higher plants. Journal of Experimental Botany 55, 2131–2145. https://doi.org/10.1093/jxb/erh248
- Tetlow, I.J., Beisel, K.G., Cameron, S., Makhmoudova, A., Liu, F., Bresolin, N.S., Wait, R., Morell, M.K., Emes, M.J., 2008. Analysis of Protein Complexes in Wheat Amyloplasts Reveals Functional Interactions among Starch Biosynthetic Enzymes. Plant Physiology 146, 1878–1891. https://doi.org/10.1104/pp.108.116244
- Vandromme, C., Spriet, C., Dauvillée, D., Courseaux, A., Putaux, J.-L., Wychowski, A., Krzewinski, F., Facon, M., D'Hulst, C., Wattebled, F., 2019. PII1: a protein involved in starch initiation that determines granule number and size in Arabidopsis chloroplast. New Phytol 221, 356–370. https://doi.org/10.1111/nph.15356
- Vandromme, C., Spriet, C., Putaux, J., Dauvillée, D., Courseaux, A., D'Hulst, C., Wattebled,
 F., 2023. Further insight into the involvement of PII1 in starch granule initiation in
 Arabidopsis leaf chloroplasts. New Phytologist 239, 132–145.
 https://doi.org/10.1111/nph.18923
- Wattebled, F., Dong, Y., Dumez, S., Delvallé, D., Planchot, V., Berbezy, P., Vyas, D., Colonna, P., Chatterjee, M., Ball, S., D'Hulst, C., 2005. Mutants of Arabidopsis Lacking a Chloroplastic Isoamylase Accumulate Phytoglycogen and an Abnormal Form of Amylopectin. Plant Physiol 138, 184–195. https://doi.org/10.1104/pp.105.059295
- Weatherwax, P., 1922. A Rare Carbohydrate in Waxy Maize. Genetics 7, 568–572.
- Wilkinson, D.L., Harrison, R.G., 1991. Predicting the Solubility of Recombinant Proteins in Escherichia coli. Nat Biotechnol 9, 443–448. https://doi.org/10.1038/nbt0591-443
- Xu, X., Dees, D., Dechesne, A., Huang, X.-F., Visser, R.G.F., Trindade, L.M., 2017. Starch phosphorylation plays an important role in starch biosynthesis. Carbohydrate Polymers 157, 1628–1637. https://doi.org/10.1016/j.carbpol.2016.11.043

- Yu, J.W., Choi, J.-S., Upadhyaya, C.P., Kwon, S.O., Gururani, M.A., Nookaraju, A., Nam, J.-H., Choi, C.-W., Kim, S.I., Ajappala, H., Kim, H.S., Jeon, J.H., Park, S.W., 2012. Dynamic proteomic profile of potato tuber during its in vitro development. Plant Science 195, 1–9. https://doi.org/10.1016/j.plantsci.2012.06.007
- Zeeman, S.C., Kossmann, J., Smith, A.M., 2010. Starch: Its Metabolism, Evolution, and Biotechnological Modification in Plants. Annu. Rev. Plant Biol. 61, 209–234. https://doi.org/10.1146/annurev-arplant-042809-112301
- Zeeman, S.C., Smith, S.M., Smith, A.M., 2007. The diurnal metabolism of leaf starch. Biochemical Journal 401, 13–28. https://doi.org/10.1042/BJ20061393
- Zeeman, S.C., Thorneycroft, D., Schupp, N., Chapple, A., Weck, M., Dunstan, H., Haldimann, P., Bechtold, N., Smith, A.M., Smith, S.M., 2004. Plastidial *a* -Glucan Phosphorylase Is Not Required for Starch Degradation in Arabidopsis Leaves But Has a Role in the Tolerance of Abiotic Stress. Plant Physiology 135, 849–858. https://doi.org/10.1104/pp.103.032631
- Zeeman, S.C., Tiessen, A., Pilling, E., Kato, K.L., Donald, A.M., Smith, A.M., 2002. Starch Synthesis in Arabidopsis. Granule Synthesis, Composition, and Structure. Plant Physiol. 129, 516–529. https://doi.org/10.1104/pp.003756
- Zhang, X., Myers, A.M., James, M.G., 2005. Mutations Affecting Starch Synthase III in Arabidopsis Alter Leaf Starch Structure and Increase the Rate of Starch Synthesis. Plant Physiology 138, 663–674. https://doi.org/10.1104/pp.105.060319
- Zhang, X., Szydlowski, N., Delvallé, D., D'Hulst, C., James, M.G., Myers, A.M., 2008. Overlapping functions of the starch synthases SSII and SSIII in amylopectin biosynthesis in Arabidopsis. BMC Plant Biol 8, 96. https://doi.org/10.1186/1471-2229-8-96



Annexe 1 : Article Bossu et al., 2024 BioRXiv

Annexe 2 : Article Osman et al., 2024

Annexe 3 : Séquence peptidique des constructions tronquées PII1 et SS4.

Annexe 4 : Alignements multiples de *At*PII1 avec les 28 orthologues identifiés sur TAIR.

1	PII1/MRC interaction with Starch Synthase 4 (SS4) from Arabidopsis thaliana
2	inhibits SS4 enzymatic activity
3	
4	Short title: PII1 inhibits SS4 enzymatic activity
5	
6	M. Bossu ¹ , R. Osman ¹ , G. Brysbaert ¹ , Marc F. Lensink ¹ , D. Dauvillée ¹ & C. Bompard ^{1*}
7	
8	¹ Univ. Lille, CNRS, UMR 8576 - UGSF - Unité de Glycobiologie Structurale et Fonctionnelle, F-
9	59000 Lille, France
10	
11	*Corresponding author
12	
13	Abstract
14	Starch is the major energy storage compound in plants. It accumulates in the form of
15	insoluble, partly crystalline granules whose number and shape are specific to each plant
16	species. These characteristics are defined very early in starch biosynthesis, at the initiation
17	stage.
18	Starch biosynthesis initiation is a complex process that relies on the coordinated action of
19	several proteins that interact together in the so-called complex of initiation. Starch Synthase
20	4 (SS4) is the only initiation protein with enzymatic activity. It catalyzes the formation of
21	glucan primers, which serve as substrates for the enzymatic machinery that synthesizes
22	starch granules. Previous studies have highlighted the importance of interactions between
23	SS4 and regulatory proteins in this process. Among them, Protein Involved in Initiation 1
24	(PII1) interacts with SS4 but its function is not yet established. In this study, we explored the
25	structural and functional implications of PII1 on SS4's enzymatic activity. Our findings reveal
26	that PII1 contains a long coiled-coil domain that specifically interacts with SS4, leading to
27	significant inhibition of SS4's glucan elongation activity. Importantly, this inhibition is specific
28	to SS4 and does not affect other known synthases, suggesting a targeted regulatory
29	mechanism. This work describes the structural specificities of PII1 and SS4 and reveals a
30	function for PII1 in the initiation complex. These results allow us to re-examine these

31 complex mechanisms and propose new hypotheses about the important steps in the32 initiation of starch biosynthesis.

33

34 Introduction

In higher plants, starch is the most abundant and widely distributed non-structural carbohydrate that accumulates as water-insoluble, partly crystalline granules. In heterotrophic tissues, starch functions as a long-term carbohydrate reserve, supporting germination or seasonal regrowth. This "storage starch" is a key component of our staple crops, accounting for half of human caloric intake and is also extensively used as an industrial commodity. In leaves, transitory starch accumulates in chloroplasts during the day and is used as carbon and energy source during the night.

42 Starch is composed of two polymers of glucose residues, namely amylose and amylopectin, 43 organized as linear α -1,4-glucans covalently linked to one another by α -1,6-bonds also called 44 branching points (for review (Pfister and Zeeman, 2016)). Amylose molecules are mainly 45 linear (<1% α -1,6 bonds) whereas amylopectin, the main component contains 5-6% of 46 branching points (Pérez and Bertoft, 2010; Pfister and Zeeman, 2016).

Amylose linear chains are synthesized by the action of the Granule Bound Starch Synthase (GBSS) that is embedded in starch granules and belongs to the starch synthase class of glucosyltransferases (Ball and Morell, 2003; Schwarte et al., 2013; Pfister and Zeeman, 2016; Seung, 2020). A second non enzymatic protein, namely PROTEIN TARGETING TO STARCH 1 (PTST1) is also involved in the process by targeting GBSS to starch granules (Seung et al., 2015).

Amylopectin is synthesized by the concerted activities of soluble starch synthases (SSs), starch branching enzymes (SBEs) and starch debranching enzymes (SDBEs) acting by forming transient complexes during the different stages of biosynthesis (Crofts et al., 2015). Soluble SSs transfer the glucose residue of ADP-Glucose (ADP-glc), the precursor molecule, to the non-reducing end of growing amylopectin and amylose molecules, or on free maltooligosaccharides (MOSs) that arise from degradation of starch granules formation (Larson et al., 2016; Seung et al., 2017; Xie et al., 2018).

Branching points are introduced by SBEs which cleave and transfer a linear maltooligosaccharide and transfer it to an α -1,6 position (Sawada et al., 2014). The isoamylase class (ISA) of SDBEs is involved in the synthesis of amylopectin by hydrolyzing the excess and incorrectly positioned α-1,6 bonds of soluble amylopectin molecule thus facilitating
amylopectin crystallization (Ball et al., 1996; Myers et al., 2000; Delatte et al., 2005;
Wattebled et al., 2005). Finally, two recently discovered non enzymatic proteins, LESV (LIKE
EARLY STARAVTION 1) and ESV1 (EARLY STARVATION 1) are involved in amylopectin packing
and starch granule protection (Liu et al., 2023; Osman et al., 2024).

68 In addition to these biosynthetic mechanisms, an initiation step is required to produce the oligosaccharide substrates of the biosynthetic enzymes. This step has been identified, but its 69 70 molecular mechanism remains poorly understood even if several proteins involved in this 71 mechanism have been identified. The first of them is an enzyme from the SSs family, starch 72 synthase 4 (SS4). Arabidopsis mutant plants lacking SS4 show a phenotype in which the number of starch granules formed in the chloroplasts is greatly reduced and the remaining 73 74 granules are much larger and spherical, in contrast to the lenticular granules observed in the 75 wild-type plants (Roldan et al., 2007). This 1040 amino acids protein is composed of three 76 distinct domains: an N-terminal domain containing a transit peptide (1-42), a predicted 77 coiled-coil region (43-465), a so-called dimerization domain (471-515), and the catalytic 78 domain at the C-terminus (540-1040). The N-terminal domain is predicted as a coiled-coil 79 (cc) domain containing four cc regions (Raynaud et al., 2016). This N-terminal domain was 80 shown to be required for the granule shape, while the catalytic domain is sufficient to regulate granule number (Lu et al., 2018). Subsequently, several non-enzymatic proteins 81 82 were identified as being involved in starch granule initiation. Arabidopsis plants that do not 83 express these proteins show a phenotype in which the number of granules is reduced, but 84 they keep their lenticular shapes (Lu et al., 2018). These proteins include two members of 85 the PTST family: PTST2 and PTST3, two conserved coiled-coil and CBM48 containing proteins 86 (Seung et al., 2017), the thylakoid-associated MAR-BINDING FILAMENT-LIKE PROTEIN 1 87 (MFP1) (Sharma et al., 2024), MYOSIN-RESEMBLING CHLOROPLAST PROTEIN or PROTEIN 88 INVOLVED IN INITIATION 1 (MRC/PII1)(Seung et al., 2018; Vandromme et al., 2019) and a 89 non-canonical starch synthase, SS5 (Abt et al., 2020).

90 It has been proposed and shown that these proteins interact specifically, probably 91 sequentially and in some cases transiently, as part of an initiation complex whose molecular 92 mechanism is still poorly understood (Seung et al., 2018). The function of some of these 93 proteins has been investigated.

94 The first event in starch initiation is likely linked to the action of the MFP1 that interacts with 95 PTST2 and specifically determines the subchloroplast location of the starch granule initiation 96 machinery (Sharma et al., 2024). PTST2 also interacts with PII1 and SS4 but interaction with 97 SS4 is not dependent of the cc region (Seung et al., 2017). Through its CBM, it binds and 98 delivers glucans, able to adopt a helical secondary structure, to SS4. The function of PTST3 is 99 partially redundant with that of PTST2 (Seung et al., 2017). The non-canonical starch 100 synthase, SS5, which contains an incomplete catalytic domain, has no enzymatic activity and 101 interacts with PII1 (Abt et al., 2020). It remains to be determined whether proteins with 102 multiple partners are able to interact with them simultaneously via multiple recognition 103 sites, or whether these interactions are in competition with each other as part of their 104 functional regulation.

105 PII1 has a predicted coiled-coil region that spans most of the protein without any predicted 106 catalytic or glucan binding domains (Seung et al., 2018). Knockout mutants have a phenotype 107 analog to ss4 mutant plants with one large lenticular (rather than spherical in ss4) granule 108 per chloroplast instead of the normal 5-7. This phenotype is similar to the *ptst2* phenotype, 109 but is less severe, as pii1 chloroplasts still contain one starch granule in most chloroplasts, 110 whereas *ptst2* and *ss4* mutants have many empty chloroplasts. PII1 function is unknown, 111 particularly the role of its interaction with SS4 in the initiation process. In this work we studied both the structure of PII1 and SS4 as well as their interaction. Through biochemical 112 113 experiments using the recombinant proteins, we were able to reveal an inhibitory effect of 114 PII1 on the elongation capacity of SS4. This unexpected result is discussed in terms of its 115 potential implication on the regulation of starch granule initiation in vivo.

116

117 Methods

118

119 Molecular Modelling

120 Protein structures of all PII1 and SS4 constructs were initially modelled using both 121 AlphaFold2 combined through MassiveFold (ref here: 122 https://www.researchsquare.com/article/rs-4319486/v1) that allows to go further by generating a large number of molecular models, all the neural network model versions 123 124 available to date (v1, v2 and v3) (Jumper and Hassabis, 2022). Initially, three predictions per 125 neural network model were generated for monomers, resulting in fifteen predictions per monomeric protein, and five predictions per neural network model for protein complexes, resulting in seventy-five predictions for each. Massive sampling and increased recycling were also attempted for some complexes, pushing predictions until six hundred, but didn't show substantially improved results in these cases, compared to a basic prediction run.

When this was available, we used AlphaFold3 for the calculation of new molecular models and compared them to the predictions already obtained by MassiveFold. The structure models obtained with both approaches being similar for our proteins, we only kept AlphaFold3 predictions for the sake of clarity of the manuscript.

With AlphaFold3, for each monomeric protein, five different predictions were computed and ranked by global pLDDT. The five molecular models generated have been superimposed and used to identify the presence of dynamic regions. Molecular models with best pLDDT and PAE values were used for figures. For each complex, fifty predictions were computed and ranked by the AlphaFold confidence score. Scores and PAE values were used to evaluate the quality of the predictions.

Structures and electronic surfaces were visualized using the PyMOL Molecular GraphicSystem, Shrödinger, LLC.

142

143 Synchrotron radiation circular dichroism

144 Synchrotron radiation circular dichroism (SR-CD) spectra were measured at the DISCO beamline of the SOLEIL Synchrotron (Gif-sur-Yvette, France). The beam size of 4 × 4 mm and 145 the photon-flux per nm step of 2×10^{10} photons s⁻¹ in the spectral band from 270–170 nm 146 147 prevented radiation-induced damage (Miles et al., 2008). CD spectra were acquired using 148 IGOR software (WaveMetrics). Before measurements the molar elliptical extinction 149 coefficient of Ammonium d-10-Camphorsulfonate Ammonium (CSA) has been measured on 150 the beamline and used as standard for calibration of all data measurements (Miles et al., 151 2004). Protein and buffer spectra were collected consecutively and are the mean of 3 152 accumulations. The buffer baseline was then subtracted from the spectra and the data 153 processing was conducted using CDToolX software (Miles and Wallace, 2018).

154 PII1-H2-H3 protein solutions were deposited between 2 CaF₂ coverslips with a pathlength of 155 10 (TFE) or 50 (protein alone) μ m (Refregiers et al., 2012). The influence of the different 156 concentration of TFE on the structure of PII1 was studied by mixing the protein and different 157 concentrations of TFE and then measuring the spectra under the same conditions as for the

native protein after a 10 min incubation. Spectra containing protein buffer at corresponding
 TFE concentrations were subtracted from the protein/TFE spectra before CSA calibration.

160 **Cloning, expression and purification of proteins**

161 PII1-H2-H3 and SS4-∆349 were cloned and expressed in *Escherichia coli* as recombinant proteins lacking their N-terminal transit peptides. The complete cDNAs encoding the 162 163 complete PII1 and SS4 lacking their N-terminal transit peptides, cloned into the pENT-D-Topo plasmid (Thermofisher, Rochester, USA)(Vandromme et al., 2019) were used as templates for 164 165 the construction of the vectors expressing the different truncated forms of SS4 and PII1. A 166 fragment of the Pii1 cDNA corresponding to the H2 and H3 helixes was amplified with 167 primers EntPII1For et EntPII1Rev using the ultra-high Kapa Hifi Hot start polymerase (Roche 168 diagnostics, Austria) and the 1.4kb PCR product was transferred into the pENT-D-Topo 169 plasmid (Thermofisher, Rochester, USA). This entry vector was used to construct the pET300-PII1 vector by using the gateway technology into the Champion[™] pET300 plasmid 170 171 (Thermofisher) that allows the generation of recombinant proteins fused to a N-terminal 6-172 Histidine tag. The PCR product corresponding to the SS4- Δ 349 truncated *ss4* cDNA fragment 173 was amplified using the Kapa Hifi Hot start polymerase and digested by BamHI and XhoI 174 restriction enzymes before its cloning into the pET-DUET-1 (Novagen) vector's corresponding 175 restriction sites allowing the expression of N-terminal His-tagged proteins. For co-expression 176 experiments, the PII1 sequence was amplified using the BamPII and SalPII primers allowing 177 the introduction of BamHI and Sall restriction sites which were used to clone the digested 178 PCR product into the pET-DUET-1. A second PCR product corresponding to the SS4- Δ 349 179 cDNA sequence was amplified using the RVSS4 and XhoSS4 primers and the PCR product was 180 cloned in the *Eco*RV and *Xho*I restriction sites of the previous plasmid allowing co-expression 181 of both N-terminal His-tagged PII1 and SS4- Δ 349 proteins.

Primers sequences are detailed in table 1, the restriction enzyme sites used for the cloning are underlined in the primers' sequences. Each construct was fully sequenced before use to ascertain the integrity of the cloned sequences.

185 The recombinant proteins were expressed in *E. coli* BL21 DE3 in 2 x 500 ml of LB medium

supplemented with 100μ g/mL antibiotic then induced with 0.5 mM IPTG overnight at 20°C.

187 Cells were pelleted at 6000 g for 30 min at 4°C and stored at -80°C for further use.

Cells pellets were resuspended in lysis buffer (20 mM Hepes/NaOH pH 7, 150 mM NaCl supplemented with one tablet per liter of EDTA-free of protease inhibitors EDTA free (Roche), disrupted by Emulsiflex or sonication and centrifuged at 10000 g for 30 min at 4°C. The supernatants were then subjected to purification.

Purification was performed by a first step of Immobilized Metal Affinity Chromatography
(IMAC) using a 5ml IMAC HisTrap Excel column (Cytiva, Amersham,UK). Protein sample was
loaded onto the column pre-equilibrated with buffer A (20 mM Hepes/NaOH pH 7, 150 mM
NaCl). The beads were washed with 10 column volume of buffer A supplemented with 10mM
Imidazole and eluted with buffer A supplemented with 500 mM Imidazole and 10% (w/v)
glycerol.

198 This initial step was followed by a second purification step through size exclusion

199 chromatography Superdex200 10/300 (Cytiva) equilibrated with buffer A. 500µl of the IMAC

eluted sample were injected on the column. The protein sample purity was assessed by SDS-PAGE 10%.

For structural study, protein samples were concentrated using Vivaspin centrifugal
 concentrator with a 10 kDa cut-off (Sartorius). Protein concentrations were determined using
 a Nanodrop Spectrophotometer (ND1000) from Thermo Scientific.

205

206 **Zymogram**

207 To study the elongation capacity of SS4, soluble proteins were separated by electrophoresis 208 onto a 10 % non-denaturing acrylamide gel containing 0.3 % oyster glycogen (Sigma). The run 209 was carried out for 2 hours at 200 mA at 4°C using the Bio-Rad Mini Protean III system. The 210 gel was then incubated overnight at room temperature in synthase incubation mix (67 mM 211 Glycyl-glycine/NaOH pH 9; 133 mM ammonium sulfate; 80 mM MgCl₂; 0.6 mg.ml¹ bovine 212 serum albumin; 25 mM β-mercaptoethanol; 2.4 mM ADP-glucose). Control experiments 213 were carried out following the same procedure but omitting the ADP-glc in the incubation 214 mix. To test the capacity of SS4 to synthesize de novo glucan primers, the same procedure 215 was used but without providing any polysaccharide substrate in the gel matrix. The initiation 216 capacity of the enzyme was tested in a control experiment in which a mix of small MOS were 217 provided in the incubation mix.

218

219 Results

220

221 Expression of PII1 and SS4

222 To investigate the role of PII1 in the initiation steps leading to new starch granules, we aimed 223 to study its structure as well as its interaction with SS4. For these studies, we considered a 224 biochemical and structural approach. As these approaches required the expression and 225 purification of both proteins, we undertook the expression of both proteins in their entirety, 226 with the exception of their chloroplast transit peptides (cTPs). To do this, we carried out a 227 series of expression assays on plasmids containing the genes encoding each protein, 228 optimized for expression in E. coli (see Methods section). We succeeded in expressing PII1 229 under these conditions, but none of the tests allowed us to obtain enough soluble protein, 230 while SS4 was little or no expressed. At this stage, and in order to continue the study, we 231 considered using truncated but functional constructs of each protein. To identify the regions 232 that might affect expression or solubility, we analyzed the sequence of PII1 and SS4 and 233 calculated molecular models using AlphaFold3.

234

235 Structural analysis of PII1 molecular model

Previous sequence analysis of PII1 predicted the presence of coiled-coil regions along almost the entire length of the protein (Seung et al., 2018; Vandromme et al., 2019) suggesting that the protein is involved in interactions with one or more other proteins. To visualize their arrangement, we then analyzed molecular models of the protein using AlphaFold3 (Abramson et al., 2024).

241 AlphaFold 3 computes five predictions and sort them by confidence score. Several criteria for 242 the validity of the structure are available, the pLDDTs, the PAE matrix and, for the multimers, 243 the TM coefficients. The pLDDTs indicates the confidence estimate per atom on a scale of 0-244 100, where a higher value indicates a higher confidence for each protein amino acid. The PAE 245 (Predicted Aligned Error) matrix estimates the error in the relative position and orientation 246 between two parts in the predicted structure. Higher values indicate higher predicted error 247 and therefore lower confidence. And then, TM scores (predicted template modeling (pTM) 248 score and the interface predicted template modeling (ipTM) score) that are both derived 249 from a measurement called the template modeling (TM) score. pTM is an integrated 250 measure of how well AlphaFold-Multimer has predicted the overall structure of the complex. 251 In contrast, ipTM measures the accuracy of the predicted relative positions of the subunits 252 forming the protein-protein complex. Disordered regions and regions with low pLDDT score may negatively impact the ipTM score even if the structure of the complex is predicted 253 254 correctly. Both measure the accuracy of the entire structure (Zhang and Skolnick, 2004; Xu 255 and Zhang, 2010), for a high confidence it should be >0.5 and >0.8 respectively. TM scores 256 should be interpreted cautiously (Abramson et al., 2024) as they can be influenced by several factors. For this study all TM scores were below the "trustable" values and this can be 257 258 attributed to the presence of large cc regions as well as predicted unfolded regions in 259 proteins. Indeed, in cc regions, local contacts can be inferred with a good score, which 260 explains the pLDDT values > 0.80 in these regions. However, the contacts can be shifted 261 between repeats in the prediction, which makes the process less straightforward, resulting in 262 low pTM scores. Disordered regions are always hard to predict, with high flexibility, resulting 263 in low pLDDT values in these regions and decreasing the pTM scores. Therefore, we 264 remained cautious in interpreting the results and models in which the PAE matrix showed no 265 credible interactions between proteins were considered invalid.

266 For the full length PII1, the five models calculated by AlphaFold3 were very similar. For the 267 sake of clarity, only the model with the best score is presented. The model obtained is 268 composed of 6 helices predicted by a majority with a pIDDT>90 confidence index and 269 numbered H1 to H6 from the N-terminus. The 70 amino acids at the N-terminus and the 270 region between helices H3 and H4 are predicted to be disordered. Helices H2 and H3 are 271 very long helices of 203 and 245 amino acids respectively, organized in antiparallel coils as 272 suggested by the sequence analysis (Figure 1A). The analysis of the PAE (Figure 1B) matrix 273 shows that the relative position of helices H1, H2 and H3 is predicted with a high confidence 274 index, which is not the case for the other helices within them, with the exception of short 275 helices H5 and H6.

The H2 and H3 helices interact with each other *via* hydrophobic interactions involving amino acids predicted to be involved in predicted repeated heptads for coiled-coils (Figure 1C). On the other hand, numerous charged residues are directed towards the protein surface and may be involved in interactions with PII1 partners. The numerous hydrophilic residues present on the surface of the H2-H3 domain of PII1, may create interactions with the solvent, increasing the chances of obtaining a soluble protein sample for its structural and functional analysis (Figure 1C). 283

284 The coiled-coil region of PII1 is conserved among orthologs

285 To verify the conservation of the coiled-coil (abbreviated cc in the rest of the manuscript) H2-286 H3 region of PII1, we investigated whether it was conserved in its orthologs. Twenty-eight 287 orthologs were identified on the Arabidopsis information resource (TAIR 288 https://www.arabidopsis.org). All the orthologs are proteins close in size to AtPII1 (between 289 720 and 820 residues), with the exception of the Setaria italica protein, which has a much 290 shorter sequence (492 residues). Their amino acid sequences were aligned using blastp 291 (Altschul et al., 1990) (Figure 2A, supplementary Figure 1).

The result shows high conserved sequence for the 27 *At*PII1 orthologs proteins analyzed, all along the sequence. *Setaria italica* has a much shorter sequence than the other orthologs, corresponding only to the N-terminal region containing H1, H2 and H3 in *At*PII1 (Figure 2A supplementary Figure 1).

As with *At*PII1, we calculated a molecular model for each of the identified PII1 orthologs using AlphaFold3. With the exception of the *Setaria* ortholog, which contains only the H1 and the coiled-coil domain, analysis of the results shows the presence of the 6 predicted helices similar to those predicted for the Arabidopsis protein (Figure 2A).

Alignment of the structures of H2 and H3 of PII1 and those of its 28 orthologs shows a high degree of structural conservation of the conserved cc region (Figure 2B), which is, in agreement with an important role for this region in the function of PII1.

303

304 Dimerization domain has a crucial role in SS4 dimer structure

The structure of the catalytic domain of *At*SS4 has been solved by crystallography in the presence of ADP-glucose and acarbose, a competitive inhibitor used in the treatment of certain types of diabetes. Analysis of the structure revealed a monomeric protein and has allowed the precise characterization of its catalytic domain (Nielsen et al., 2018).

Four cc regions (cc1 to cc4 numbered from the N-terminus) have been identified in the Nterminal part of SS4 (Raynaud et al., 2016) followed by a so-called dimerization domain upstream the catalytic domain. In this work, SS4 has been shown to be dimeric and that it is in this form that it is the most active. The dimerization is linked to the presence of the dimerization domain, and cc regions are not involved in. The cc1 and cc2 regions are required for SS4 localization to thylakoid. Interestingly, the authors produced a truncated 315 construct of the cc1 and cc2 regions (comprising the first 349 amino acids of the protein) 316 which proved to be soluble (Raynaud et al., 2016) and, in order to obtained soluble protein 317 samples, we decided to work with the same construct, which we will call SS4- Δ 349, for the 318 remainder of this work.

The dimerization domain is predicted to be located between amino acids 471 and 515 and its structure is not known (Raynaud et al., 2016). In order to identify this domain, to verify the possibility of dimerization and to visualize the position of the different regions that make up SS4, we created dimer models of SS4 and SS4- Δ 349 using AlphaFold3.

323 The two models obtained show the regions containing the predicted catalytic and 324 dimerization domains with a high level of confidence (pLDDT above 90 and between 70 and 325 90 respectively) and the analysis of the PAE matrix shows a good level of confidence in the 326 relative positions of these domains (Supplementary Figure 2A, B, C, D). The AlphaFold3 327 predicted catalytic domain of SS4 is almost identical of the X-ray-crystallographic structure 328 (Nielsen et al., 2018) with a rmsd of 0.3 Å for all C α . On the predicted structure, SS4 has a 329 predicted dimeric structure in which the dimerization domains of each monomer interact 330 with the GT5 subunit of the catalytic domain of the second monomer sharing a large surface 331 area (Figure 3A). The pattern of the dimerization domain, originally predicted to be between 332 residues 471 and 515, is composed of residues 466 to 533 in the model. Its structure is 333 predicted as a globular domain composed of a bundle of 4 helices (dH1 to dH4) parallel to 334 each other and an N-terminal region of about 7 amino acids forming a short helix (Figure 3B). 335 The two dimerization domains are organized into swap domains, each of which interacts with 336 the GT5 domain of the neighboring molecule (Figure 3A). This interaction involves the helices 337 dH1 and dH2 of the dimerization domain and the GT5 domain of the adjacent molecule 338 described in Figure 3B, D. The interaction between the dimerization domain and the GT5 339 domain involves numerous hydrophilic and hydrophobic interactions between the side and 340 main chains of residues of about ten amino acids on each side. The interaction zone covers a 341 large surface area of both domains, indicating a strong interaction. This interaction between 342 dimerization domain and GT5 is conserved in complete or truncated structure of SS4 (Figure3) which is a further vote of confidence in the already very convincing pLDDT and PAE 343 344 matrix values (Supplementary Figure 2 C, D). In the SS4 full length model, interactions are 345 predicted between the two dimerization domains by the short N-terminal helix 346 (supplementary Figure 2E). In the truncated model SS4- Δ 349, the interaction between the

347 dimerization domains and the GT5 domains is identical to the full SS4 model (rmsd 0,144 Å for all C α), with the same confidence indices and positioning errors of the same order 348 349 (Supplementary Figure 2C, Figure 3D). In contrast, the interaction between the dimerization 350 domains observed in the SS4 is not conserved in the model of SS4- Δ 349 in which the two 351 dimerization domains are not in the same position and are not directly interacting 352 (supplementary Figure 2E, F). This difference in the predictions is unlikely due to the full or 353 truncated version of the dimer to predict, but rather to the uncertainty in the positioning of 354 the dimerization domains between each other, as shown in both PAE matrix plots. This could indicate that, in the context of conformational changes associated either with the 355 356 binding/release of glucans or with interaction with another protein in the initiation complex, 357 the interaction between the dimerization domain of one monomer and the GT5 domain of 358 the adjacent monomer would remain stable, whereas the interaction between the two 359 dimerization domains could change.

360

361 N-terminal domain of SS4 organizes in coiled-coil

The N-terminal domain, although predicted as helices for the cc1 to cc4 regions, have different configurations depending on whether the whole protein or the truncated protein is considered (and on the position of the dimeric domains in relation to each other).

365 In the five SS4 models generated by AlphaFold3 the structure of the catalytic and 366 dimerization domains is conserved in contrast to the N-terminal domain. The latter is always 367 predicted to be organized as a coiled-coil, but the distribution of helices, predicted with a 368 lower confidence index than the C-terminal domain, is guite different. In the best model, as 369 expected, the 1-180 region is predicted to be disordered, while the cc1/cc2 and cc3/cc4 regions are involved in two distinct 200 Å long helices. The cc3/cc4 helices of each monomer 370 371 are predicted with good confidence (pLDDT between 90 and 70) whereas the cc1/cc2 helices 372 are predicted with low confidence (pLDDT <70) (supplementary Figure2 A to D). Altogether, in the model they form a tetrameric coiled-coil confirmed by the PAE. In this organization, 373 374 the two cc3/cc4 regions form a central dimeric coiled-coil around which the cc1 and cc2 regions are supercoiled (supplementary Figure3 A). It seems likely that AlphaFold positioned 375 376 cc1 and cc2 on cc3/cc4 to stabilize the latter. cc1 and cc2 have been shown to be involved in 377 (and essential for) the localization of SS4 to thylakoids. In this case, they must form a domain distinct from cc3/cc4 in order to interact with MFP1 or PTST2. It is therefore highly likely thatthey are not organized as in the model described.

380 The five SS4- Δ 349 AlphaFold models were also very similar. In these models the cc3 and cc4 381 regions form distinct helices forming an antiparallel cc that assembles with the cc3/cc4 region of the second molecule, whose folding is identically predicted, to form a confident 382 tetrameric cc (90> pLDDT >70 and 70> pLDDT >50) folded with low expected position error 383 384 (supplementary Figure 2C, D, supplementary Figure 3B). This configuration is probably 385 adopted to stabilize the hydrophobic regions of cc3/cc4 that interact with cc1/cc2 in the 386 model of the whole protein. Like for the full-length protein, this could indicate that individual 387 cc regions are able to adapt their folding depending on the partner with which they interact. 388 This would give SS4 the ability to adapt to multiple partners (individually or together) by 389 adjusting the conformation of its cc regions.

390

391 SS5 has a similar dimeric organization as SS4

392 Since SS5 is a protein involved in the initiation mechanisms and interacts with PII1, we 393 studied its structure using AlphaFold3 (Abramson et al., 2024). SS5, like SS4, is predicted to 394 be dimeric (Abt et al., 2020), so we started by modelling a dimer. The model obtained has a 395 high confidence index, as shown by the pLDDTs per residue and the PAE matrix 396 (supplementary Figure 4A, B).

The first 59 amino acids are predicted to be disordered and residues 60-139 of the two molecules form two 115 Å long helices forming a coiled-coil as observed for SS4. These helices are longer than the previous predicted cc region in SS5 (residues 68 to 120) (Abt et al., 2020).

401 The molecular model shows a dimeric organization of SS5, similar to that of SS4, involving 402 two dimerization swapping-domains (from residue 121 to 139), one of each monomer (supplementary Figure 4C). The dimerization domain of SS4 is not conserved in SS5 sequence 403 404 (Abt et al., 2020), however a dimerization domain is observed with a conserved structure 405 when compared with SS4(supplementary Figure 4D). The helix H1 which interacts with the 406 GT5 domain of the other monomer is involved in the predicted long N-terminal helices 407 forming the coiled-coil region. It has been shown that when the previous predicted cc region 408 of SS5 (residues 68-120) is absent, the protein is neither able to form dimers nor to interact with PII1 (Abt et al., 2020). This result is consistent with the molecular model of SS5, with theabsence of the cc region likely to affect the structure of the dimerization domain.

The GT5 domain of SS5 is extended by a helix (439-460) positioned at the site occupied by the C-terminal helix originating from the GT1 domain of SS4 (which is absent in SS5) and interacting with the GT5 domain of SS5 (supplementary Figure 4E).

The model for SS5 is very similar, in terms of its N-terminal, cc, dimerization and GT5 domains, with the addition of the terminal helix, to SS4 at the dimerization site. The GT5 domain, the dimerization domain and the Helix 439-460 of SS5 superpose to the equivalent structures in SS4 with a rmsd of 1 Å on 250 C α (supplementary Figure 4E). Only these two SSs have this dimeric organization and both interact with PII1, so it is conceivable that the dimerization region could be involved in the interaction with PII1.

420

421 PII1 H2-H3 is soluble and organized in CC

To study PII1 and SS4, we made truncated constructs to obtain soluble samples compatible with biochemical and biophysical approaches. For PII1, we produced a protein construct containing only the H2-H3 helices that make up the conserved cc region, hereafter referred to as PII1-H2-H3.

426 PII1-H2-H3 is expressed in large amounts, most of which is found in the soluble fraction 427 (Supplementary Figure 5). The soluble fraction was incubated with IMAC beads and eluted 428 with 500 mM imidazole. At this stage, and despite incubation with the beads, only part of the 429 soluble sample was able to bind to the resin and eluted with imidazole. The fraction that did 430 not bind was subjected to a second unsuccessful incubation with Ni-beads, suggesting that in 431 solution the protein must adopt multiple conformations or aggregate and that in this form 432 the 6-histidine tag is no longer accessible. The eluted fraction was then subjected to a steric 433 exclusion chromatography (SEC) step. PII1-H2-H3 eluted in a very broad peak, probably due 434 to its non-globular structure, and the presence of the long cc. This type of chromatography is 435 probably not suitable for non-globular proteins and does not allow us to determine the 436 monodispersity of the sample, but it did allow us to eliminate the contaminants present at 437 the end of the affinity chromatography step. We therefore kept it in the protein purification 438 protocol (Figure 4A).

To experimentally verify the coiled-coil structure of PII1-H2-H3, we used a Synchrotron radiation circular dichroism (SR-CD) study approach on the purified PII1-H2-H3 protein 441 sample at 3 mg/ml. This method is mostly used to analyze the content of secondary 442 structure elements in a protein and to distinguish between individual helices and helices 443 involved in a coiled-coil assembly, which is particularly relevant to our study. A typical 444 spectrum obtained from alpha helices shows a positive peak around 190 nm and two 445 negative peaks at 208 and 222 nm, the latter mainly due to helices, the others also coming 446 from other structural elements. For free helices the magnitude of the 208 nm pic is lower 447 than the one at 222 nm, the opposite is true for helices involved in a coiled-coil (Wallace, 448 2009). SR-CD analysis of PII1-H2-H3 presents a typical helix-folded protein profile (Figure 4B) 449 with a positive peak at 190 nm and two negative peaks at 208 and 222 nm. The ratio of the 450 ellipticity measured at 208 nm to that measured at 222 nm gives a value lower than 1, indicating and confirming the organization of the helices into cc. 451

452 To go further in the validation of the presence of cc regions, we measured different SR-CD 453 spectra of PII1-H2-H3 in the presence of increasing amounts of 2,2,2-trifluoroethanol (TFE), 454 which has the property of separating helices organized in cc and stabilize them as individual 455 helices. The results show a rather atypical effect of TFE on PII1-H2-H3 at low concentrations 456 of TFE (10%), the protein helices are initially unfolded, causing the spectrum to flatten. At 20-457 30% TFE, the helices gradually reform, resulting in a maximum signal corresponding to 458 individualized helices at 50% TFE concentrations in which all helices are independent (Θ 222 459 $/\Theta$ 208>1) (Figure 4C) thus definitely confirming the structuration of helices in cc in the 460 structure of PII1. The fact that PII1 can easily unfold and then reform helices is guite atypical 461 and shows that the protein is highly labile under certain conditions, allowing it to undergo 462 major conformational changes.

463

464 **SS4-**∆**349** is soluble and displays enzymatic activity

Since SS4 cannot be produced in its full form, we decided to express a truncated version of the protein for the biochemical assays. We based this on the work of (Raynaud et al., 2016), who expressed a soluble and active SS4 construct containing the cc3, cc4, dimerization and catalytic domains. The deleted part corresponds to the first 349 residues, so we named this construct SS4- Δ 349.

470 SS4- Δ 349 was expressed in small quantity but in soluble form. It was purified by an IMAC 471 affinity chromatography step (Figure 5A). 472 We then verified that deletion of cc1/cc2 in SS4 did not affect the ability of the enzyme to 473 elongate glucan chains. To do this, we performed zymograms consisting of the separation by 474 electrophoresis of soluble protein extracts of bacteria expressing SS4- Δ 349 (150µg of soluble 475 proteins), under native conditions, in a 10% polyacrylamide gel containing 0.3% glycogen. 476 After migration we observed the capacity of the enzyme to elongate the outer chains of 477 glycogen as revealed after iodine staining of the gels after incubation in the presence of ADP-478 Glc. This activity is revealed by the appearance of a brown/black coloring band with iodine, 479 while the unmodified glycogen chains stains orange/light brown. To verify that the measured 480 activity was not due to a bacterial enzyme, we performed the same experiment with a 481 soluble extract of the same bacterial strain that does not express SS4- Δ 349. (Figure 5B). 482 Analysis of the results showed very low elongation activity for the BL21 strain likely 483 corresponding to the E. coli glycogen synthase (GlgA) at the top of the gel. In contrast, a band 484 of high elongation activity was observed in the strain expressing SS4- Δ 349, indicating the 485 high activity of the enzyme in its truncated form. This activity is shared by most of the starch 486 synthases isoforms targeted to the chloroplast. It is clearly not this polyglucan elongation 487 activity, even if it exists, that gives SS4 such an important role in plants, but its initiation 488 activity. Studies using high performance anion exchange chromatography coupled to pulsed 489 amperometric detection (HPAEC-PAD) showed that all SSs are capable of extending glucan 490 chains from ADP-glucose and maltose or glucans with a higher degree of polymerization (DP) 491 (Brust et al., 2013). None of them, including SS4, could synthesize de novo glucans from ADP-492 glucose and glucose (Brust et al., 2013). We tried to perform the same experiment by 493 replacing glycogen with a mix of small MOS, but we were unable to detect any iodine 494 staining.

495

496 **PII1 inhibits specifically SS4 activity**

In order to check if the interaction with PII1 had an effect on the glucan elongating activity of SS4, we mixed soluble extracts from bacteria in which PII1-H2-H3 and SS4- Δ 349 had been expressed. Since we could not measure the exact amount of each protein in the soluble extract, we first mixed an equivalent amount of total protein for each extract (175 µg and 350 µg, Figure 6A). After incubation for 15 minutes, the mixtures were loaded onto a polyacrylamide gel and the SS4 elongation activity was assessed as previously by iodine staining of the zymogram after incubation overnight with ADP-Glc (Figure 6A).

504 The ratio of each protein was estimated from the intensity of the PII1-H2H3 and SS4- Δ 349 505 bands on SDS-PAGE. (Supplementary Figure 6A). This analysis showed that in the soluble 506 extracts the amount of PII1 was about 60 times higher than that of SS4. This value should be 507 treated with caution as the level of expression (especially for SS4) varies from one expression 508 to another. For this experiment, we performed several controls in order to analyze the results 509 unambiguously. First, we performed the same experiment without ADP-glucose to check that 510 the observed activity was not linked to another substrate (Figure 6). We then performed the 511 experiment with soluble proteins from expression bacteria that were not transformed by the 512 plasmid to check that the observed activity was due to SS4- Δ 349 and not to a bacterial protein (Figure 6A). We did the same with bacteria expressing PII1-H2-H3 to verify that this 513 514 protein doesn't have any enzymatic activity (Figure 6A). Finally, we checked that the activity 515 of SS4- Δ 349 remains unaffected in the presence of an excess amount of BSA (Figure 6B). To 516 verify the specificity of the effect of PII1-H2-H3 on SS4- Δ 349, we compared it with the effect 517 on GlgA activity and tested the effect of increasing amounts of PII1 (Supplementary Figure 518 6B).

519 Analysis of the results shows that no elongating activity is observed in the absence of ADP-520 glucose, confirming that this precursor is used as a substrate by SS4- Δ 349 and GlgA and that 521 no effect of the presence of BSA could be observed on the activity of both SS4- Δ 349 and 522 GlgA (Figure 6B). PII1-H2-H3 doesn't display any elongation activity.

In the presence of PII1-H2-H3, a strong inhibition of SS4- Δ 349 activity is observed under the tested conditions (Figure 6A, Supplementary Figure 6B), and this inhibition is proportional to the amount of PII1-H2-H3 in the reaction mixture. However, the presence of PII1-H2-H3 does not affect the activity of bacterial GlgA as SS4- Δ 349 loses its elongation activity whereas GlgA remains fully active regardless of the amount of PII1-H2-H3. This demonstrates the specificity of PII1-H2-H3's inhibitory effect on SS4.

529 This experiment shows that PII1-H2-H3 interacts specifically to SS4- Δ 349 and that this 530 interaction inhibits SS4- Δ 349 glucan elongation activity.

531

532 Molecular modelling of the interaction between SS4 proteins involved in the initiation 533 complex

534 Numerous interactions between the proteins involved in initiation have been demonstrated. 535 PII1 can interact strongly with SS4, PTST2 and SS5 (Seung et al., 2018; Abt et al., 2020). PTST2 536 interacts strongly with PTST3 and MFP1 weaker with PII1 and transiently with the C-terminal 537 domain of SS4 through its CBM (Seung 2017,2018). PTST3 interacts only with PTST2. To 538 visualize the interaction modes of these proteins, we attempted to model all possible 539 combinations of interaction using AlphaFold3 (Abramson et al., 2024). We calculated 50 540 structure models with AlphaFold3 for each complex and analyzed the results very carefully. 541 Unfortunately, we were unable to obtain any models that met our selection criteria. In most 542 cases we obtained models with reasonable pLDDT values, of the same order as the models 543 obtained for the individual proteins, but the PAE matrices did not validate the positions of 544 each molecule. Furthermore, of the 50 structure models calculated, most proposed different 545 structures, sometimes even clashes between the different chains. The ipTM and pTM values 546 never exceeded 0.35 and 0.4 respectively. This is probably due to several factors: most of the 547 proteins involved in the initiation complex (all except the catalytic domain of SS4) have 548 sections with very few homologous sequences, especially in the cc regions, regions with 549 repeats which can be organized in a large number of structural variations, which makes the 550 training and subsequent inference harder, especially when several partners are involved, 551 knowing the variety of binding they can be involved in (see discussion), and all the proteins 552 have disordered regions which reduce the prediction scores of the obtained models

553

554 Discussion

All the proteins involved in starch binding initiation have been proposed to interact with each other to form an initiation complex (Seung et al., 2018; Abt et al., 2020). This proposal was supported by interaction studies between the identified molecules and by the fact that all these proteins have more or less important regions of their structure predicted to be organized in coiled-coils, known to be involved in protein-protein interactions (Truebestein and Leonard, 2016).

As their physical properties, particularly their length and flexibility, have important structural and functional properties, the cc are widely found in biological processes and multimolecular complexes (for review see (Truebestein and Leonard, 2016). Long cc, act often as molecular spacers, either separating functional domains or scaffolding large macromolecular complexes. Some cc act simply as molecular spacers or facilitating oligomerization, while others have evolved the ability to communicate conformational changes along their length.
Some of them are able to bind membranes, link two compartments in the cell (Engel et al.,
1992) conferring conformational plasticity. The coiled-coil domain of myosin, which is similar
to PII1, has been shown to act as a mechanosensor and to transmit structural changes along
its chain (Linari et al., 2015).

571 At the present stage of our knowledge of initiation mechanisms, we know that all the 572 proteins involved have one or more cc regions predicted on the basis of sequence, and we 573 have information on interactions between proteins that have been shown experimentally. 574 However, apart from the crystallographic structure of the catalytic domain of SS4 and the 575 CBM48 domains of the PTSTs, there is no structural information and very little knowledge 576 about the regions involved in the interactions. Nor do we know whether the initiation 577 complex involves all the proteins at once, or whether the interactions are transient or 578 sequential.

579 In the present work we focused on the structural study of PII1 and SS4 combined with 580 structural and biochemical approaches to determine the effect of the binding of PII1 on SS4 581 enzymatic activity.

582

583 PII1 contains a long conserved coiled-coil domain

584 We studied the structure of PII1 using molecular modelling and showed that the protein is 585 composed of 6 helices that are conserved in all identified orthologs found on the TAIR 586 website except Setaria italica, which has a shorter sequence corresponding only to the H1 H2 587 and H3 helices of AtPII1. The conserved sequences also adopt a conserved 3D structure, in 588 particular the long H2-H3 helices, which account for almost 60% of the amino acids in the 589 protein. These two helices join together to form a long cc region, which is about 300-350 Å in 590 length. The presence of cc was demonstrated experimentally by SR-CD, but could not be 591 observed by SAXS due to the conformational heterogeneity of the solution. We have shown 592 that this domain is conserved in all tested orthologs and that it may play an important role in 593 the function of PII1.

At this stage, it is difficult to envisage a function for PII1 solely on the basis of its structure. However, it has been shown that PII1 interacts with PTST2, SS5 and SS4, and it may be involved in regulating SS4 activity.

597

598 What is the advantage of organizing SS4 and SS5 into dimers?

599 Both SS4 and SS5 form dimers (Raynaud et al., 2016; Abt et al., 2020). We showed that a 600 structurally conserved dimerization region located between the N-terminal cc region and the 601 C-terminal catalytic domain is involved in the dimer formation. Our results show with high 602 confidence that these dimerization regions form a small globular domain that interacts with 603 the GT5 subunit of the SS4 catalytic domain or the SS5 C-terminal domain.

604 This organization, in which each monomer exchanges its dimerization domain, interacting 605 strongly with the GT5 domain of the second monomer, offers some interesting properties. 606 Firstly, it allows the two GT5 domains of the catalytic domains to be kept close together, 607 leaving the catalytic region (for SS4) or the glucan-binding domain (for SS5) accessible. On 608 the other hand, the dimerization domain forms a hinge region between the GT5 domains 609 and the cc regions, the structure of which confers dynamic properties that allow signals to be 610 transmitted from the cc domain to the catalytic domain via conformational changes induced 611 by the binding of other partners. It is therefore conceivable that the binding of PTST2 and/or 612 PII1 to SS4 could influence its catalytic activity in the initiation complex.

SS4 and SS5 both interact specifically with PII1 which does not recognize other families ofSSs.

615 All SSs share a catalytic domain (for functional SSs) or at least the GT5 domain and nonconserved N-terminus domains. SS1 and SS2 have no predicted cc regions, in contrast to SS3, 616 617 SS4 and SS5 (Abt et al., 2020). Only SS4 and SS5 interact with PII1. The specificity of their 618 recognition by PII1 cannot be entirely related to the presence of cc regions, since SS3 is not 619 recognized. Nor can it be due to the structure of GT5, which is present in all SSs. What 620 distinguishes SS4 and SS5 from other SSs is the dimerization associated with the presence of 621 a structurally conserved dimerization domain that interacts strongly with the GT5 domain 622 and acts as a hinge with the cc region. Since this dimerization is necessary for recognition by 623 PII1 (Abt et al., 2020), it is conceivable that the latter interacts with SS4 and SS5 specifically 624 with this part of the molecule and (due to the structure of PII1) with the coiled-coil regions 625 of the SSs. Dimerization could therefore be the prerogative of SSs involved in the initiation of 626 starch biosynthesis. This also implies that PII1 interacts with either SS4 or SS5, but not both 627 at the same time. In this case, SS5 could have a regulatory effect on the binding of PII1 to 628 SS4.

630 **PII1 can inhibit SS4 enzymatic activity.**

631 Apart from the fact that it interacts with SS4, PTST2 and SS5 and that its absence in the plant results in a phenotype in which the initiation of starch granule synthesis is affected, the role 632 633 of PII1 is not known. In this paper, we have shown that PII1 conserved cc interacts with a 634 truncated form of SS4 and that this interaction inhibits the elongating activity of SS4. We 635 have also shown that this inhibition is specific to SS4 and does not affect the activity of the 636 *E.coli* glycogen synthase GlgA. Since GlgA and SS4 have similar catalytic domains, this result 637 indicates that PII1 probably does not interact directly with the catalytic domain of SS4, but 638 rather via the cc domain of the enzyme (which is also part of the dimerization) and/or its 639 dimerization region.

640 It is rather counterintuitive to observe the inhibition of SS4 by PII1, since the absence of PII1 641 in the plant leads to a reduction in initiation events. This inhibition can only occur transiently 642 in a complex, highly dynamic mechanism in which interactions within the initiation complex 643 are transient or evolving. We have therefore re-examined the current data in the light of our 644 results.

645 As shown by the structural analysis, the dimerization domain of SS4 forms a hinge region 646 between the cc region and the catalytic domain. It is likely that binding of PII1 to SS4 induces 647 a conformational change in the catalytic domain leading to its loss of activity maybe braking 648 the link between the two catalytic domains or separating both GT domain. However, the 649 strength of this inhibition must be qualified in the absence of other initiation proteins, in 650 particular PTST2, which binds to the catalytic domain of SS4 via its CBM48 domain (Seung et 651 al., 2017). PTST2 interacts as well with MFP1 and this interaction is involved in the 652 localization of SS4 to thylakoid membranes (Sharma et al., 2024). The sites of interaction 653 between PTST2, PII1 and MFP1 are not known, but since the latter two are almost entirely 654 cc-structured, it is likely that this interaction occurs via the cc region of PTST2. As PII1 has 655 conserved helices (outside the cc domain) and PII1 and PTST2 have predicted structurated 656 and unstructured regions, we cannot rule out that these regions are involved in the 657 interaction, increasing the possibility of signal transduction by conformational change. Our 658 circular dichroism results showed that PII1 could easily reorganize into other structures, 659 indicating a certain lability in its structuring.

The CBM48 domain of PTST2 binds specifically helically folded glucans with DP 10 or higherthat are the shortest MOS able to form crystalline glucans. It has been proposed that PTST2

662 brings these substrates to SS4 to induce initiation (Seung et al., 2017). However, in a yeast 663 model system expressing different combinations of starch biosynthetic enzymes, the 664 presence of SS4 tended to promote the formation of insoluble glucans, and this effect was 665 seen in the absence of PTST2 (Pfister and Zeeman, 2016). These glucans can be used as substrate by SS1 for amylopectin biosynthesis. If such substrates exist in the chloroplast, why 666 667 would an initiation step be required if SS1 can use them? On the other hand, PTST2 is 668 required for the initiation of biosynthesis and is also responsible for poly-initiation when 669 overexpressed in plants. Another hypothesis for its function may be that helical glucans 670 would not be substrates but products of the reaction catalyzed by SS4. The function of PTST2 671 may be to receive these products and make them available to the enzymes involved in starch biosynthesis. In that way PII1, which inhibits SS4 activity, could intervene to stop the 672 673 elongation of glucan chains once they are sufficiently long to emerge from the catalytic 674 region, at which point they would be retrieved by PTST2. Once released, the polyglucan 675 could be taken up by the enzymatic biosynthetic machinery and PTST2 could reposition itself 676 to suppress inhibition by PII1 allowing then SS4 to initiate a new primer. Our hypothesis is 677 that in the absence of PII1, SS4 initiates the synthesis of the glucan chain, but since the 678 enzyme can no longer be inhibited by PII1, it behaves like the other SSs and forms much 679 longer chains, reducing the number of primers available and therefore the number of 680 initiation events. SS5 may play a role in regulating PII1 function.

681 In conclusion, the fact that each of these molecules has one or more cc regions, giving them 682 the ability to interact with each other, but also to transmit conformational change signals 683 within the initiation complex itself, suggests a complex and highly regulated mechanism of 684 which we currently know only the basics. Our current knowledge is based on analyses of the 685 phenotypes of mutants in which the initiation proteins have been inactivated or 686 overexpressed, and on identification of partners, which shed light on the nature of the 687 interactions but do not allow us to go any further. In this work, we have studied the 688 structural and enzymatic aspects of the current state of knowledge, which has allowed us to 689 make progress in our understanding of the mechanisms and also to put forward some new 690 hypotheses. The difficulty of producing the key players in the initiation complex in solution, 691 and the presence of long cc regions that tend to aggregate in the absence of partners, 692 complicates conventional structural biology tools that require monodisperse solutions. 693 Determining the stoichiometry of each complex formed in solution and using cryo-electron-

694 microscopy to study them will undoubtedly be an asset in the study of these complex 695 mechanisms. The development of AI tools for modelling molecular models of multi-protein 696 complexes, which is rapidly expanding, should soon make it possible to predict the 697 interactions between the different initiation proteins and greatly improve our understanding 698 of these mechanisms.

699

Table 1: List of oligonucleotide primers used in this study. The restriction enzyme sites areunderlined.

702

Oligonucleotide	DNA Sequence 5' -3'	Amplicon sizes	Encoded sequence
Primer name		(bp)	
EntPII1For	CACCCTGCCATACTCGAATCTTGGC	1396	102-564
EntPII1Rev	TTAGCGGTTAGTGAGCTCTGCAATACG		PII1 H2 and H3 helices
Bam∆349For	CCC <u>GGATCC</u> CGAATGCACTGATCTATGG	2094	350-1040
XhoSS4Rev	CC <u>CTCGAG</u> TTACGTGCGATTAGGAACAGCTC		
BamPII	GG <u>GGATCC</u> CGTGAATCATAAACAAAAAGC		
SalPII	GGG <u>GTCGAC</u> TTACAGGATAACGCCTTCCATAC	2250	Full length PII1 devoid of TP
RVSS4∆349	CCC <u>GGATCC</u> CGAATGCACTGATCTATGG		
XhoSS4∆349	CC <u>CTCGAG</u> TTACGTGCGATTAGGAACAGCTC	2094	350-1040

703

704 Bibliography

706	Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L,
707	Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung CC, O'Neill M, Reiman D,
708	Tunyasuvunakool K, Wu Z, Zemgulyte A, Arvaniti E, Beattie C, Bertolli O, Bridgland
709	A, Cherepanov A, Congreve M, Cowen-Rivers AI, Cowie A, Figurnov M, Fuchs FB,
710	Gladman H, Jain R, Khan YA, Low CMR, Perlin K, Potapenko A, Savy P, Singh S,
711	Stecula A, Thillaisundaram A, Tong C, Yakneen S, Zhong ED, Zielinski M, Zidek A,
712	Bapst V, Kohli P, Jaderberg M, Hassabis D, Jumper JM (2024) Accurate structure
713	prediction of biomolecular interactions with AlphaFold 3. Nature
714	Abt MR, Pfister B, Sharma M, Eicke S, Burgy L, Neale I, Seung D, Zeeman SC (2020) STARCH
715	SYNTHASE5, a Noncanonical Starch Synthase-Like Protein, Promotes Starch Granule
716	Initiation in Arabidopsis. Plant Cell 32: 2543-2565
717	Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool.
718	J Mol Biol 215: 403-410
719	Ball S, Guan HP, James M, Myers A, Keeling P, Mouille G, Buleon A, Colonna P, Preiss J
720	(1996) From glycogen to amylopectin: a model for the biogenesis of the plant starch
721	granule. Cell 86: 349-352
722	Ball SG, Morell MK (2003) From bacterial glycogen to starch: understanding the biogenesis
723	of the plant starch granule. Annu Rev Plant Biol 54: 207-233

724	Brust H, Orzechowski S, Fettke J, Steup M (2013) Starch Synthesizing Reactions and Paths: in
725	vitro and in vivo Studies. J. Appl. Glycosci. 60: 3-20
726	Crofts N, Abe N, Oitome NF, Matsushima R, Hayashi M, Tetlow IJ, Emes MJ, Nakamura Y,
727	Fujita N (2015) Amylopectin biosynthetic enzymes from developing rice seed form
728	enzymatically active protein complexes. J Exp Bot 66: 4469-4482
729	Delatte T, Trevisan M, Parker ML, Zeeman SC (2005) Arabidopsis mutants Atisa1 and Atisa2
730	have identical phenotypes and lack the same multimeric isoamylase, which influences
731	the branch point distribution of amylopectin during starch synthesis. Plant J 41: 815-
732	830
733	Engel AM, Cejka Z, Lupas A, Lottspeich F, Baumeister W (1992) Isolation and cloning of Omp
734	alpha, a coiled-coil protein spanning the periplasmic space of the ancestral
735	eubacterium Thermotoga maritima. EMBO J 11: 4369-4378
736	Larson ME, Falconer DJ, Myers AM, Barb AW (2016) Direct Characterization of the Maize
737	Starch Synthase IIa Product Shows Maltodextrin Elongation Occurs at the Non-
738	reducing End. J Biol Chem 291: 24951-24960
739	Linari M, Brunello E, Reconditi M, Fusi L, Caremani M, Narayanan T, Piazzesi G, Lombardi V,
740	Irving M (2015) Force generation by skeletal muscle is controlled by mechanosensing
741	in myosin filaments. Nature 528: 276-279
742	Liu C, Pfister B, Osman R, Ritter M, Heutinck A, Sharma M, Eicke S, Fischer-Stettler M,
743	Seung D, Bompard C, Abt MR, Zeeman SC (2023) LIKE EARLY STARVATION 1 and
744	EARLY STARVATION 1 promote and stabilize amylopectin phase transition in starch
745	biosynthesis. Sci Adv 9: eadg7448
746	Lu KJ, Pfister B, Jenny C, Eicke S, Zeeman SC (2018) Distinct Functions of STARCH SYNTHASE
747	4 Domains in Starch Granule Formation. Plant Physiol 176 : 566-581
748	Miles AJ, Janes RW, Brown A, Clarke DT, Sutherland JC, Tao Y, Wallace BA, Hoffmann SV
749	(2008) Light flux density threshold at which protein denaturation is induced by
750	synchrotron radiation circular dichroism beamlines. J Synchrotron Radiat 15: 420-422
751	Miles AJ. Wallace BA (2018) CDtoolX. a downloadable software package for processing and
752	analyses of circular dichroism spectroscopic data. Protein Sci 27: 1717-1722
753	Miles AJ, Wien F, Wallace BA (2004) Redetermination of the extinction coefficient of
754	camphor-10-sulfonic acid, a calibration standard for circular dichroism spectroscopy.
755	Anal Biochem 335: 338-339
756	Myers AM, Morell MK, James MG, Ball SG (2000) Recent progress toward understanding
757	biosynthesis of the amylopectin crystal. Plant Physiol 122 : 989-997
758	Nielsen MM, Ruzanski C, Krucewicz K, Striebeck A, Cenci U, Ball SG, Palcic MM, Cuesta-
759	Seijo JA (2018) Crystal Structures of the Catalytic Domain of Arabidopsis thaliana
760	Starch Synthase IV, of Granule Bound Starch Synthase From CLg1 and of Granule
761	Bound Starch Synthase I of Cyanophora paradoxa Illustrate Substrate Recognition in
762	Starch Synthases. Front Plant Sci 9: 1138
763	Osman R, Bossu M, Dauvillee D, Spriet C, Liu C, Zeeman SC, D'Hulst C, Bompard C (2024)
764	LIKE EARLY STARVATION 1 interacts with amylopectin during starch biosynthesis. Plant
765	Physiol 195: 1851-1865
766	Pérez S, Bertoft E (2010) The molecular structures of starch components and their
767	contribution to the architecture of starch granules: A comprehensive review Starch-
768	Stärke 62: 389-420
769	Pfister B, Zeeman SC (2016) Formation of starch in plant cells. Cell Mol Life Sci 73: 2781-2807

- Raynaud S, Ragel P, Rojas T, Merida A (2016) The N-terminal Part of Arabidopsis thaliana
 Starch Synthase 4 Determines the Localization and Activity of the Enzyme. J Biol
 Chem 291: 10759-10771
- Refregiers M, Wien F, Ta HP, Premvardhan L, Bac S, Jamme F, Rouam V, Lagarde B, Polack F,
 Giorgetta JL, Ricaud JP, Bordessoule M, Giuliani A (2012) DISCO synchrotron radiation circular-dichroism endstation at SOLEIL. J Synchrotron Radiat 19: 831-835
- Roldan I, Wattebled F, Mercedes Lucas M, Delvalle D, Planchot V, Jimenez S, Perez R, Ball S,
- D'Hulst C, Merida A (2007) The phenotype of soluble starch synthase IV defective
 mutants of Arabidopsis thaliana suggests a novel function of elongation enzymes in
 the control of starch granule formation. Plant J 49: 492-504
- Sawada T, Nakamura Y, Ohdan T, Saitoh A, Francisco PB, Jr., Suzuki E, Fujita N, Shimonaga T,
 Fujiwara S, Tsuzuki M, Colleoni C, Ball S (2014) Diversity of reaction characteristics of
 glucan branching enzymes and the fine structure of alpha-glucan from various
 sources. Arch Biochem Biophys 562: 9-21
- Schwarte S, Brust H, Steup M, Tiedemann R (2013) Intraspecific sequence variation and
 differential expression in starch synthase genes of Arabidopsis thaliana. BMC Res
 Notes 6: 84
- 787 Seung D (2020) Amylose in starch: towards an understanding of biosynthesis, structure and
 788 function. New Phytol 228: 1490-1504
- Seung D, Boudet J, Monroe J, Schreier TB, David LC, Abt M, Lu KJ, Zanella M, Zeeman SC
 (2017) Homologs of PROTEIN TARGETING TO STARCH Control Starch Granule Initiation
 in Arabidopsis Leaves. Plant Cell 29: 1657-1677
- Seung D, Schreier TB, Burgy L, Eicke S, Zeeman SC (2018) Two Plastidial Coiled-Coil Proteins
 Are Essential for Normal Starch Granule Initiation in Arabidopsis. Plant Cell 30: 1523 1542
- Seung D, Soyk S, Coiro M, Maier BA, Eicke S, Zeeman SC (2015) PROTEIN TARGETING TO
 STARCH is required for localising GRANULE-BOUND STARCH SYNTHASE to starch
 granules and for normal amylose synthesis in Arabidopsis. PLoS Biol 13: e1002080
- Sharma M, Abt MR, Eicke S, Ilse TE, Liu C, Lucas MS, Pfister B, Zeeman SC (2024) MFP1
 defines the subchloroplast location of starch granule initiation. Proc Natl Acad Sci U S
 A 121: e2309666121
- 801 Truebestein L, Leonard TA (2016) Coiled-coils: The long and short of it. Bioessays **38**: 903-916
- Vandromme C, Spriet C, Dauvillee D, Courseaux A, Putaux JL, Wychowski A, Krzewinski F,
 Facon M, D'Hulst C, Wattebled F (2019) PII1: a protein involved in starch initiation
 that determines granule number and size in Arabidopsis chloroplast. New Phytol 221:
 356-370
- Wallace BA (2009) Protein characterisation by synchrotron radiation circular dichroism
 spectroscopy. Q Rev Biophys 42: 317-370
- Wattebled F, Dong Y, Dumez S, Delvalle D, Planchot V, Berbezy P, Vyas D, Colonna P,
 Chatterjee M, Ball S, D'Hulst C (2005) Mutants of Arabidopsis lacking a chloroplastic
 isoamylase accumulate phytoglycogen and an abnormal form of amylopectin. Plant
 Physiol 138: 184-195
- Xie Y, Barb AW, Hennen-Bierwagen TA, Myers AM (2018) Direct Determination of the Site of
 Addition of Glucosyl Units to Maltooligosaccharide Acceptors Catalyzed by Maize
 Starch Synthase I. Front Plant Sci 9: 1252
- 815
- 816 Acknowledgements and funding

817 The authors strongly acknowledge DISCO beamline at synchrotron SOLEIL (St Aubin, France)

and are grateful for the expert technical support provided by beamline staffs. The authors

819 would like to thank Dr. Fabrice Wattebled for providing plasmids expressing full-length PII1

and SS4 and helpful discussions. This work has been supported by Centre National de la

Recherche Scientifique France (cora funding to G.B, M.L., D.D., C.B. and PhD grant for R.O.),
Université de Lille France (PhD grant to M.B), and Région Hauts de France (PhD grant for

Université de Lille France (PhD grant to M.B), and Région Hauts de France (PhD grant forM.B.).

824

825 Author Contributions

CB, MB and DD conceived and designed the experiments. MB performed experiments with
input from CB, RO, DD, ML and GB. CB, RO and MB collected synchrotron data. MB and CB
analyzed data with the input of DD. CB wrote the manuscript. CB, MB DD, GB and ML revised
the manuscript.

830

831 Figure Legends

832

833 Figure 1: Molecular structure of PII1. The structures are represented in cartoon A) AlphaFold3 model of PII1, regions with pLDDT > 90 are colored navy blue, regions with 90 834 835 >pLDDT > 70 are colored cyan, regions with pLDDT < 70 are colored yellow and orange. The 836 six helices of the model are labeled H1 to H6 from the N-terminal end of the protein. B) PAE 837 matrix of the model, the region corresponding to the six helices are labeled. C) potential 838 surface of the coiled-coil region (H2-h3) of PII1, the positive and negative charges at the 839 surface are colored blue and red respectively. The hydrophobic side chains of amino acids involved in the coiled-coil stabilization are shown in black sticks. 840

841

842 Figure 2: Primary and tertiary structure alignment of AtPII1 with 28 orthologs A) Amino 843 acid sequence alignment of PII1 with the 27 orthologs having the same length computed 844 with Blastp. The sequence of Setaria italica has been manually added from the alignment of 845 PII1 with all tested orthologs. Both alignments are detailed in Supplementary Figure 1. Only 846 alignment positions with no gaps are colored. Red indicates highly conserved positions and 847 blue indicates lower conservation, gaps are represented by red lines and grey regions 848 correspond to non-conserved regions. The position of conserved helices is indicated by 849 arrows and labeled below B) Structure alignment of H2-H3 coiled-coil region of PII1 with the 850 28 orthologs. Structures are represented in cartoon, helices and loops are colored red and 851 green respectively.

852

853 Figure 3: Molecular models of dimeric SS4 and SS4-∆349. The structures are represented in 854 cartoon A) structural organization of SS4 full length as predicted by AlphaFold 3. Only one 855 monomer is colored (catalytic domain in red, dimerization domain in green, cc3-4 helix in 856 violet and cc1-2 helix in orange). The catalytic sites are indicated by a black star. The second 857 monomer is colored gray. B) molecular model of the dimerization domain of SS4. The 858 positions of helices are labeled dH1 to dH4 from the N-terminal. C) structural organization of 859 SS4- Δ 349 with the same color code as SS4 D) Interaction between the dimerization (in red) 860 and GT5 (in green) domains of SS4. The side chains of amino-acids residues involved in the 861 interaction are represented in sticks.

- Figure 4: Purification and SR-CD analysis of PII1-H2-H3 A) SDS-PAGE of the protein sample of
 PII1-H2-H3 after purification, used for SR-CD experiments. B) SR-CD spectrum of PII1-H2-H3
 C) SR-CD spectra of PII1-H2-H3 at different concentrations of TFE. The color code used for
 spectra in function of TFE concentration are indicated.
- Figure 5: Purification and characterization of starch synthase elongation activity of SS4-A349. A) SDS-PAGE of the protein sample of SS4- Δ 349 after purification B) Zymogram revealing the elongation function of SS4- Δ 349. 150 µg of total soluble protein extract of BL21 (DE3) and BL21 (DE3) expressing SS4- Δ 349 were loaded onto a 10 % non-denaturing acrylamide gel containing 0.3 % oyster glycogen (Sigma). After migration, the gel was incubated overnight at room temperature in synthase incubation buffer containing 2.4 mM ADP-glucose. Starch synthase activities were visualized as brown bands after iodine staining.
- 876 Figure 6: Zymogram revealing PII1-H2-H3 effect on SS4- Δ 349 starch synthase activity. A) 877 175 or 350 µg of total soluble protein extract of BL21 (DE3) and BL21 (DE3) expressing SS4-878 Δ 349 or PII1-H2-H3 were loaded onto a 10 % non-denaturing acrylamide gel containing 0.3 % 879 oyster glycogen (Sigma). After migration, the gel on the left was incubated overnight at room 880 temperature in synthase incubation buffer containing 2.4 mM ADP-glucose and the gel on 881 the right in synthase incubation buffer without ADP-glucose. Starch synthase activities were 882 visualized as brown bands after iodine staining. B) Zymogram analyzing the effect of BSA on 883 SS4- Δ 349 activity.
- 884

867

- Supplementary Figure 1: Primary structure alignment of AtPii1 with plant orthologs A)
 Sequence alignment of AtPII1 with all the analyzed orthologs B) sequence alignment of
 AtPII1 orthologs except ortholog from Seratia italica
- 888
- **Supplementary Figure 2: AlphaFold models of dimeric SS4 and SS4-** Δ **349** A) structure of SS4 B) and SS4- Δ 349. regions with pLDDT > 90 are colored navy blue, regions with 90 >pLDDT > 70 are colored cyan, regions with 70 >pLDDT > are colored yellow and 50 pLDDT < 50 are colored orange. C) PAE matrix for SS4 D) and SS4- Δ 349. E) Dimeric organization of SS4 and F) SS4- Δ 349. Only the catalytic and dimerization domains of each monomer are represented. One monomer is represented as green cartoon and the second is represented as red molecular surface.
- 896

Supplementary Figure 3: Structural analysis of the N-terminal tetracoil regions of SS4 and SS4- Δ 349 dimers A) In SS4, the two cc3/cc4 helices forming a central dimeric coiled-coil are represented as surface potential and the cc1-cc2 helices that are supercoiled around are represented as cartoon. B) In SS4- Δ 349 dimers, the cc3 and cc4 regions form distinct helices forming an antiparallel cc that assembles with the cc3/cc4 region of the second molecule. The cc3 and CC4 helices of one monomer are represented in surface potential whereas the cc3 and cc4 helices of the second monomer are represented in cartoon.

904

Supplementary Figure 4: AlphaFold models of SS5 dimers A) Structure of SS5, regions with
 pLDDT > 90 are colored navy blue, regions with 90 >pLDDT > 70 are colored cyan, regions
 with 70 >pLDDT > are colored yellow and 50 pLDDT < 50 are colored orange. B) PAE matrix
 for SS5 dimer C) structural organization of SS5. Only one monomer is colored (GT5 domain in
 red, dimerization domain in green, the helix (439-460 in cyan). The second monomer is

- 910 colored gray **D)** Dimeric organization of SS5. Only the GT5 and dimerization domains of each
- 911 monomer are represented. One monomer is represented as green cartoon and the second is
- 912 represented as red molecular surface. **E)** Structure superposition of the GT5 and dimerization
- 913 domains of SS5 (in purple) and SS4 (in green for GT5 and orange for the dimerization domain
- 914 and the C-terminal helix of GT1).
- 915
- 916 Supplementary Figure 5: SDS-PAGE showing expression and solubility of PII1-H2-H3 and
- 917 **SS4-\Delta349**. Crude bacterial extracts before induction (0), after induction (I) and soluble
- 918 fractions (S) have been loaded on a 10% SDS-PAGE. Migration bands of PII1-H2-H3 and SS4-
- 919 $$\Delta349$$ are indicated by green and black arrows respectively.
- 920
- 921 Supplementary Figure 6: Influence of SS4- Δ 350 / PII1-H2-H3 ratio on SS4 elongation
- 922 activity. A) SDS-PAGE loaded with different quantities (in μ g) of SS4- Δ 350 (in black) and PII1-
- 923 H2-H3. (in green) bacterial soluble extract. B) Zymogram revealing the effect of increasing
- 924 amounts of PII1-H2-H3 on SS4- Δ 349 starch synthase activity.



Figure 1: Molecular structure of PII1. The structures are represented in cartoon A) AlphaFold3 model of PII1, regions with pLDDT > 90 are colored navy blue, regions with 90 >pLDDT > 70 are colored cyan, , regions with pLDDT < 70 are colored yellow and orange. The six helices of the model are labeled H1 to H6 from the N-terminal end of the protein. B) PAE matrix of the model, the region corresponding to the six helices are labeled. C) potential surface of the coiled-coiled region (H2-h3) of PII1, the positive and negative charges at the surface are colored blue and red respectively. The hydrophobic side chains of amino acids involved in the coiled-coil stabilization are shown in black sticks.

Figure 1



Figure 2: Primary and tertiary structure alignment of AtPII1 with 28 orthologs A) Amino acid sequence alignment of PII1 with the 27 orthologs having the same length computed with Blastp. The sequence of *Setaria italica* has been manually added from the alignment of PII1 with all tested orthologs. Both alignments are detailed in Supplementary Figure 1. Only alignment positions with no gaps are colored. Red indicates highly conserved positions and blue indicates lower conservation, gaps are represented by red lines and grey regions correspond to non-conserved regions. The position of conserved helices is indicated by arrows and labeled below B) Structure alignment of H2-H3 coiled-coil region of PII1 with the 28 orthologs. Structures are represented in cartoon, helices and loops are colored red and green respectively.



Figure 2

0

B



Figure 3 : Molecular models of dimeric SS4 and SS4- Δ 349. The structures are represented in cartoon A) structural organization of SS4 full length as predicted by AlphaFold 3. Only one monomer is colored (catalytic domain in red, dimerization domain in green, cc3-4 helix in violet and cc1-2 helix in orange). The catalytic sites are indicated by a black star. The second monomer is colored gray. B) molecular model of the dimerization domain of SS4. The position of helices are labeled dH1 to dH4 from the N-terminal. C) structural organization of SS4- Δ 349 with the same color code as SS4 D) Interaction between the dimerization (in red) and GT5 (in green) domains of SS4. The side chains of amino-acids residues involved in the interaction are represented in sticks.


Figure 4: Purification and SR-CD analysis of PII1-H2-H3 **A)** SDS-PAGE of the protein sample of PII1-H2-H3 after purification, used for SR-CD experiments. **B)** SR-CD spectrum of PII1-H2-H3 **C)** SR-CD spectra of PII1-H2-H3 at different concentrations of TFE. The color code used for spectra in function of TFE concentration are indicated.

Figure 4



Figure 5: Purification and characterization of starch synthase activity of SS4- Δ 349. A) SDS-PAGE of the protein sample of SS4- Δ 349 after purification B) Zymmogram revealing the elongation function of SS4- Δ 349. 150 µg of total soluble protein extract of BL21 (DE3) and BL21 (DE3) expressing SS4- Δ 349 were loaded onto a 10 % non-denaturing acrylamide gel containing 0.3 % oyster glycogen (Sigma). After migration, the gel was incubated overnight at room temperature in synthase incubation buffer containing 2.4 mM ADP-glucose. Starch synthase activities were visualized as brown bands after iodine staining.



Figure 6: Zymmogram revealing PII1-H2-H3 effect on SS4- Δ 349 starch synthase activity. A) 175 or 350 µg of total soluble protein extract of BL21 (DE3) and BL21 (DE3) expressing SS4- Δ 349 or PII1-H2-H3 were loaded onto a 10 % non-denaturing acrylamide gel containing 0.3 % oyster glycogen (Sigma). After migration, the gel on the left was incubated overnight at room temperature in synthase incubation buffer containing 2.4 mM ADP-glucose and the gel on the right in synthase incubation buffer without ADP-glucose. Starch synthase activities were visualized as brown bands after iodine staining. B) Zymmogram analyzing the effect of BSA on SS4- Δ 349 activity.

Figure 6



Supplementary Figure 1:

Primary structure alignment of AtPII1 with plant orthologs A) Sequence alignment of AtPII1 with all the analysed orthologs B) sequence alignment of AtPII1 orthologs except ortholog from *Seratia italica*



Supplementary Figure 2: AlphaFold models of dimeric SS4 and SS4- Δ 349 A) structure of SS4 B) and SS4- Δ 349. regions with pLDDT > 90 are colored navy blue, regions with 90 >pLDDT > 70 are colored cyan, regions with 70 >pLDDT > are colored yellow and 50 pLDDT < 50 are colored orange. C) PAE matrix for SS4 D) and SS4- Δ 349. E) Dimeric organization of SS4 and F) SS4- Δ 349. Only the catalytic and dimerization domains of each monomer are represented. One monomer is represented as green cartoon and the second is represented as red molecular surface.



Supplementary Figure 3: Structural analysis of the N-terminal tetracoil regions of SS4 and SS4- Δ 349 dimers A) In SS4, the two cc3/cc4 helices forming a central dimeric coiled-coil are represented as surface potential and the cc1-cc2 helices that are supercoiled around are represented as cartoon. B) In SS4- Δ 349 dimers, the cc3 and cc4 regions form distinct helices forming an antiparallel cc that assembles with the cc3/cc4 region of the second molecule. The cc3 and CC4 helices of one monomer are represented in surface potential whereas the cc3 and cc4 helices of the second monomer are represented in cartoon.



Supplementary Figure 4: AlphaFold models of SS5 dimers A) structure of SS5, regions with pLDDT > 90 are colored navy blue, regions with 90 > pLDDT > 70 are colored cyan, regions with 70 > pLDDT > are colored yellow and 50 pLDDT < 50 are coloured orange. B) PAE matrix for SS5 dimer C) structural organization of SS5. Only one monomer is colored (GT5 domain in red, dimerization domain in green, the helix (439-460 in cyan). The second monomer is coloured gray. D) Dimeric organization of SS5. Only the GT5 and dimerization domains of each monomer are represented. One monomer is represented as green cartoon and the second is represented as red molecular surface. E) Structure superposition of the GT5 and dimerization domains of SS5 (in purple) and SS4 (in green for GT5 and orange for the dimerization domain and the C-terminal helix of GT1)



Supplementary Figure 5: SDS-PAGE showing expression and solubility of PII1-H2-H3 and SS4- Δ 349. Crude bacterial extracts before induction (0), after induction (I) and soluble fractions (S) have been loaded on a 10% SDS-PAGE. Migration bands of PII1-H2-H3 and SS4- Δ 34 are indicated by green and black arrows respectively



Supplementary Figure 6: Influence of SS4- Δ 350 / PII1-H2-H3 ratio on SS4 elongation activity A) SDS-PAGE loaded with different quantities (in µg) of SS4- Δ 350 (in black) and PII1-H2-H3 (in green) bacterial soluble extract. B) Zymmogram revealing the effect of increasing amounts of PII1-H2-H3 on SS4- Δ 349 starch synthase activity.

Parsed Citations

Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung CC, O'Neill M, Reiman D, Tunyasuvunakool K, Wu Z, Zemgulyte A, Arvaniti E, Beattie C, Bertolli O, Bridgland A, Cherepanov A, Congreve M, Cowen-Rivers AI, Cowie A, Figurnov M, Fuchs FB, Gladman H, Jain R, Khan YA, Low CMR, Perlin K, Potapenko A, Savy P, Singh S, Stecula A, Thillaisundaram A, Tong C, Yakneen S, Zhong ED, Zielinski M, Zidek A, Bapst V, Kohli P, Jaderberg M, Hassabis D, Jumper JM (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature Google Scholar: Author Only Title Only Author and Title

Abt MR, Pfister B, Sharma M, Eicke S, Burgy L, Neale I, Seung D, Zeeman SC (2020) STARCH SYNTHASE5, a Noncanonical Starch Synthase-Like Protein, Promotes Starch Granule Initiation in Arabidopsis. Plant Cell 32: 2543-2565 Google Scholar: <u>Author Only Author and Title</u>

Attschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410 Google Scholar: <u>Author Only Title Only Author and Title</u>

Ball S, Guan HP, James M, Myers A, Keeling P, Mouille G, Buleon A, Colonna P, Preiss J (1996) From glycogen to amylopectin: a model for the biogenesis of the plant starch granule. Cell 86: 349-352

Google Scholar: Author Only Title Only Author and Title

Ball SG, Morell MK (2003) From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. Annu Rev Plant Biol 54: 207-233

Google Scholar: Author Only Title Only Author and Title

Brust H, Orzechowski S, Fettke J, Steup M (2013) Starch Synthesizing Reactions and Paths: in vitro and in vivo Studies. J. Appl. Glycosci. 60: 3-20

Google Scholar: Author Only Title Only Author and Title

Crofts N, Abe N, Oitome NF, Matsushima R, Hayashi M, Tetlow IJ, Emes MJ, Nakamura Y, Fujita N (2015) Amylopectin biosynthetic enzymes from developing rice seed form enzymatically active protein complexes. J Exp Bot 66: 4469-4482 Google Scholar: Author Only Title Only Author and Title

Delatte T, Trevisan M, Parker ML, Zeeman SC (2005) Arabidopsis mutants Atisa1 and Atisa2 have identical phenotypes and lack the same multimeric isoamylase, which influences the branch point distribution of amylopectin during starch synthesis. Plant J 41: 815-830

Google Scholar: Author Only Title Only Author and Title

Engel AM, Cejka Z, Lupas A, Lottspeich F, Baumeister W (1992) Isolation and cloning of Omp alpha, a coiled-coil protein spanning the periplasmic space of the ancestral eubacterium Thermotoga maritima. EMBO J 11: 4369-4378 Google Scholar: Author Only Title Only Author and Title

Larson ME, Falconer DJ, Myers AM, Barb AW (2016) Direct Characterization of the Maize Starch Synthase IIa Product Shows Maltodextrin Elongation Occurs at the Non-reducing End. J Biol Chem 291: 24951-24960 Google Scholar: Author Only Title Only Author and Title

Linari M, Brunello E, Reconditi M, Fusi L, Caremani M, Narayanan T, Piazzesi G, Lombardi V, Irving M (2015) Force generation by skeletal muscle is controlled by mechanosensing in myosin filaments. Nature 528: 276-279 Google Scholar: Author Only Title Only Author and Title

Liu C, Pfister B, Osman R, Ritter M, Heutinck A, Sharma M, Eicke S, Fischer-Stettler M, Seung D, Bompard C, Abt MR, Zeeman SC (2023) LIKE EARLY STARVATION 1 and EARLY STARVATION 1 promote and stabilize amylopectin phase transition in starch biosynthesis. Sci Adv 9: eadg7448

Google Scholar: Author Only Title Only Author and Title

Lu KJ, Pfister B, Jenny C, Eicke S, Zeeman SC (2018) Distinct Functions of STARCH SYNTHASE 4 Domains in Starch Granule Formation. Plant Physiol 176: 566-581

Google Scholar: Author Only Title Only Author and Title

Miles AJ, Janes RW, Brown A, Clarke DT, Sutherland JC, Tao Y, Wallace BA, Hoffmann SV (2008) Light flux density threshold at which protein denaturation is induced by synchrotron radiation circular dichroism beamlines. J Synchrotron Radiat 15: 420-422 Google Scholar: <u>Author Only Title Only Author and Title</u>

Miles AJ, Wallace BA (2018) CDtoolX, a downloadable software package for processing and analyses of circular dichroism spectroscopic data. Protein Sci 27: 1717-1722

Google Scholar: <u>Author Only Title Only Author and Title</u>

Miles AJ, Wien F, Wallace BA (2004) Redetermination of the extinction coefficient of camphor-10-sulfonic acid, a calibration standard for circular dichroism spectroscopy. Anal Biochem 335: 338-339

Google Scholar: Author Only Title Only Author and Title

Myers AM, Morell MK, James MG, Ball SG (2000) Recent progress toward understanding biosynthesis of the amylopectin crystal. Plant Physiol 122: 989-997

Google Scholar: Author Only Title Only Author and Title

Nielsen MM, Ruzanski C, Krucewicz K, Striebeck A, Cenci U, Ball SG, Palcic MM, Cuesta-Seijo JA (2018) Crystal Structures of the Catalytic Domain of Arabidopsis thaliana Starch Synthase IV, of Granule Bound Starch Synthase From CLg1 and of Granule Bound Starch Synthase I of Cyanophora paradoxa Illustrate Substrate Recognition in Starch Synthases. Front Plant Sci 9: 1138 Google Scholar: <u>Author Only Title Only Author and Title</u>

Osman R, Bossu M, Dauvillee D, Spriet C, Liu C, Zeeman SC, D'Hulst C, Bompard C (2024) LIKE EARLY STARVATION 1 interacts with amylopectin during starch biosynthesis. Plant Physiol 195: 1851-1865 Google Scholar: <u>Author Only Title Only Author and Title</u>

Pérez S, Bertoft E (2010) The molecular structures of starch components and their contribution to the architecture of starch granules: A comprehensive review. . Starch-Stärke 62: 389-420

Google Scholar: Author Only Title Only Author and Title

Pfister B, Zeeman SC (2016) Formation of starch in plant cells. Cell Mol Life Sci 73: 2781-2807 Google Scholar: <u>Author Only Title Only Author and Title</u>

Raynaud S, Ragel P, Rojas T, Merida A (2016) The N-terminal Part of Arabidopsis thaliana Starch Synthase 4 Determines the Localization and Activity of the Enzyme. J Biol Chem 291: 10759-10771

Google Scholar: <u>Author Only Title Only Author and Title</u>

Refregiers M, Wien F, Ta HP, Premvardhan L, Bac S, Jamme F, Rouam V, Lagarde B, Polack F, Giorgetta JL, Ricaud JP, Bordessoule M, Giuliani A (2012) DISCO synchrotron-radiation circular-dichroism endstation at SOLEIL. J Synchrotron Radiat 19: 831-835

Google Scholar: Author Only Title Only Author and Title

Roldan I, Wattebled F, Mercedes Lucas M, Delvalle D, Planchot V, Jimenez S, Perez R, Ball S, D'Hulst C, Merida A (2007) The phenotype of soluble starch synthase IV defective mutants of Arabidopsis thaliana suggests a novel function of elongation enzymes in the control of starch granule formation. Plant J 49: 492-504

Google Scholar: Author Only Title Only Author and Title

Sawada T, Nakamura Y, Ohdan T, Saitoh A, Francisco PB, Jr., Suzuki E, Fujita N, Shimonaga T, Fujiwara S, Tsuzuki M, Colleoni C, Ball S (2014) Diversity of reaction characteristics of glucan branching enzymes and the fine structure of alpha-glucan from various sources. Arch Biochem Biophys 562: 9-21

Google Scholar: Author Only Title Only Author and Title

Schwarte S, Brust H, Steup M, Tiedemann R (2013) Intraspecific sequence variation and differential expression in starch synthase genes of Arabidopsis thaliana. BMC Res Notes 6: 84

Google Scholar: Author Only Title Only Author and Title

Seung D (2020) Amylose in starch: towards an understanding of biosynthesis, structure and function. New Phytol 228: 1490-1504 Google Scholar: <u>Author Only Title Only Author and Title</u>

Seung D, Boudet J, Monroe J, Schreier TB, David LC, Abt M, Lu KJ, Zanella M, Zeeman SC (2017) Homologs of PROTEIN TARGETING TO STARCH Control Starch Granule Initiation in Arabidopsis Leaves. Plant Cell 29: 1657-1677 Google Scholar: Author Only Title Only Author and Title

Seung D, Schreier TB, Burgy L, Eicke S, Zeeman SC (2018) Two Plastidial Coiled-Coil Proteins Are Essential for Normal Starch Granule Initiation in Arabidopsis. Plant Cell 30: 1523-1542

Google Scholar: <u>Author Only Title Only Author and Title</u>

Seung D, Soyk S, Coiro M, Maier BA, Eicke S, Zeeman SC (2015) PROTEIN TARGETING TO STARCH is required for localising GRANULE-BOUND STARCH SYNTHASE to starch granules and for normal amylose synthesis in Arabidopsis. PLoS Biol 13: e1002080

Google Scholar: Author Only Title Only Author and Title

Sharma M, Abt MR, Eicke S, Ilse TE, Liu C, Lucas MS, Pfister B, Zeeman SC (2024) MFP1 defines the subchloroplast location of starch granule initiation. Proc Natl Acad Sci U S A 121: e2309666121 Google Scholar: Author Only Title Only Author and Title

Truebestein L, Leonard TA (2016) Coiled-coils: The long and short of it. Bioessays 38: 903-916 Google Scholar: <u>Author Only Title Only Author and Title</u>

Vandromme C, Spriet C, Dauvillee D, Courseaux A, Putaux JL, Wychowski A, Krzewinski F, Facon M, D'Hulst C, Wattebled F (2019) PII1: a protein involved in starch initiation that determines granule number and size in Arabidopsis chloroplast. New Phytol 221: 356-370

Google Scholar: <u>Author Only Author and Title</u>

Wallace BA (2009) Protein characterisation by synchrotron radiation circular dichroism spectroscopy. Q Rev Biophys 42: 317-370 Google Scholar: <u>Author Only Title Only Author and Title</u>

Wattebled F, Dong Y, Dumez S, Delvalle D, Planchot V, Berbezy P, Vyas D, Colonna P, Chatterjee M, Ball S, D'Hulst C (2005) Mutants of Arabidopsis lacking a chloroplastic isoamylase accumulate phytoglycogen and an abnormal form of amylopectin. Plant Physiol 138: 184-195

Google Scholar: Author Only Title Only Author and Title

Xie Y, Barb AW, Hennen-Bierwagen TA, Myers AM (2018) Direct Determination of the Site of Addition of Glucosyl Units to Maltooligosaccharide Acceptors Catalyzed by Maize Starch Synthase I. Front Plant Sci 9: 1252 Google Scholar: Author Only Title Only Author and Title

Acknowledgements and funding

The authors strongly acknowledge DISCO beamline at synchrotron SOLEIL (St Aubin, France) and are grateful for the expert technical support provided by beamline staffs. The authors would like to thank Dr. Fabrice Wattebled for providing plasmids expressing full-length PII1 and SS4 and helpful discussions. This work has been supported by Centre National de la Recherche Scientifique France (cora funding to G.B, M.L., D.D., C.B. and PhD grant for R.O.), Université de Lille France (PhD grant to M.B), and Région Hauts de France (PhD grant for M.B.).

Author Contributions

CB, MB and DD conceived and designed the experiments. MB performed experiments with input from CB, RO, DD, ML and GB. CB, RO and MB collected synchrotron data. MB and CB analyzed data with the input of DD. CB wrote the manuscript. CB, MB DD, GB and ML revised the manuscript.

Plant Physiology®

LIKE EARLY STARVATION 1 interacts with amylopectin during starch biosynthesis

Rayan Osman (),¹ Mélanie Bossu (),¹ David Dauvillée (),¹ Corentin Spriet (),^{1,2} Chun Liu (),³ Samuel C. Zeeman (),³ Christophe D'Hulst (),¹ Coralie Bompard (),¹*

1 Université de Lille, CNRS, UMR 8576—UGSF—Unité de Glycobiologie Structurale et Fonctionnelle, Lille, France

2 University Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, US 41—UAR 2014—PLBS, Lille F-59000, France

3 Institute of Molecular Plant Biology, ETH Zurich, Universitätstrasse 2, 8092 Zurich

*Author for correspondence: coralie.bompard@univ-lille.fr

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/General-Instructions) is Coralie Bompard.

Abstract

Starch is the major energy storage compound in plants. Both transient starch and long-lasting storage starch accumulate in the form of insoluble, partly crystalline granules. The structure of these granules is related to the structure of the branched polymer amylopectin: linear chains of glucose units organized in double helices that align to form semicrystalline lamellae, with branching points located in amorphous regions between them. EARLY STARVATION 1 (ESV1) and LIKE EARLY STARVATION 1 (LESV) proteins are involved in the maintenance of starch granule structure and in the phase transition of amylopectin, respectively, in Arabidopsis (*Arabidopsis thaliana*). These proteins contain a conserved tryptophan-rich C-terminal domain folded into an antiparallel β -sheet, likely responsible for binding of the proteins to starch, and different N-terminal domains whose structure and function are unknown. In this work, we combined biochemical and biophysical approaches to analyze the structures of LESV and ESV1 and their interactions with the different starch polyglucans. We determined that both proteins interact with amylopectin but not with amylose and that only LESV is capable of interacting with amylopectin during starch biosynthesis. While the C-terminal domain interacts with amylopectin in its semicrystalline form, the N-terminal domain of LESV undergoes induced conformational changes that are probably involved in its specific function of mediating glucan phase transition. These results clarify the specific mechanism of action of these 2 proteins in the biosynthesis of starch granules.

Introduction

In plants, starch accumulates as water–insoluble, partly crystalline granules. In leaves, transitory starch accumulates in chloroplasts during the day and is used as carbon and energy source during the night. In heterotrophic tissues, storage starch accumulates over longer time frames and fuels germination or seasonal regrowth. Starch granules are made up of 2 polymers of glucose residues, namely amylose and amylopectin, that adopt different 3-dimensional (3D) structures (for review (Pfister and Zeeman 2016)) and have different physicochemical properties. They are organized as α -1,4-glucans linked one to another by α -1,6-bonds (= branching points). The major polyglucan of starch is amylopectin (70% to 80% of the starch content) which is organized by alternating regions containing linear chains or branching points, while amylose is poorly branched (<1% α -1,6 bonds) (Pérez and Bertoft 2010; Pfister and Zeeman 2016).

Amylopectin is synthesized by the concerted activities of soluble starch synthases (SSs), starch branching enzymes (SBEs), and starch debranching enzymes (SDBEs) acting independently or in concert, in particular by forming transient complexes during the various stages of biosynthesis (Crofts et al. 2015). Soluble SSs transfer the glucose residue of Downloaded from https://academic.oup.com/plphys/article/195/3/1851/7640812 by CNRS user on 01 July 2024

Received January 02, 2024. Accepted March 13, 2024. Advance access publication April 4, 2024.

[©] The Author(s) 2024. Published by Oxford University Press on behalf of American Society of Plant Biologists. All rights reserved. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

ADP-glucose (the precursor molecule) to the nonreducing end of an elongating glucan (Larson et al. 2016; Xie et al. 2018). Branching points are introduced by SBEs which cleave an α -1,4 bond of a preexisting glucan and transfer the malto-oligosaccharide located toward the nonreducing end onto a neighboring glucan or onto another part of the cleaved glucan, forming an α -1,6 bond (Sawada et al. 2014). The isoamylase class (ISA1) of DBEs are involved in the synthesis of amylopectin by hydrolyzing the excess and incorrectly positioned α -1,6 bonds of the nascent, soluble amylopectin molecule to optimize the branching pattern and facilitate amylopectin crystallization (Ball et al. 1996; Myers et al. 2000; Delatte et al. 2005; Wattebled et al. 2005). The inactivation of ISA1 in the plant induces the accumulation of phytoglycogen (an abnormal soluble glucan socalled because of some similarity to glycogen structure) and alters the structure of the residual insoluble starch, which depends on the pattern of branching points and linear chains (Zeeman et al. 1998; Delatte et al. 2005; Wattebled et al. 2005; Pfister et al. 2014). Prior to the discovery of LESV and while certain forms of insoluble altered amylopectin are still present in the isa1 mutant plants, it was widely assumed that the starch granule matrix formation involves selforganization physical process events during the early stages of granule formation (Waight et al. 2000; Ziegler et al. 2005) following the action of ISA1 (Zeeman et al. 1998; Delatte et al. 2005; Wattebled et al. 2005; Streb et al. 2008).

Within amylopectin, the entwining of adjacent chains into double helices gives rise to both secondary and tertiary structures of the molecules. These structures align and pack into dense, crystalline lamellae, which alternate with amorphous lamellae that contain most branching points and chain segments connecting the crystalline lamellae. The resulting regular alternating pattern of crystalline and amorphous layers is a feature of all wild-type starches and is believed to underlie the frequently observed 9 to 10 nm repeat structure (Buleon et al. 1998).

EARLY STARVATION 1 (ESV1) and LIKE EARLY STARVATION 1 (LESV) have been described to be involved in starch granule stabilization and phase transition of amylopectin molecules from soluble to crystalline form in starch granules, respectively (Feike et al. 2016; Liu et al. 2023). ESV1 and LESV were discovered in starch from Arabidopsis leaves and potato (Solanum tuberosum) tuber, but the genes are conserved across the plant kingdom and the orthologous proteins were present in starches of cassava, maize, and rice (Feike et al. 2016; Helle et al. 2018). The implication of ESV1 and LESV in starch metabolism has been demonstrated after the analysis of KO mutant lines of Arabidopsis in which the starch phenotype has been specifically altered (Feike et al. 2016; Liu et al. 2023). Mutant esv1 plants show a phenotype in which the diel cycle of transitory starch metabolism is altered and carbon reserves are abnormally exhausted too early before dawn. In contrast, mutant lesv plants accumulate phytoglycogen beside insoluble starch granules, especially when initiating starch synthesis de novo. Complementation

experiments have shown that overexpression of LESV in ISA1-deficient plants or yeast (*Saccharomyces cerevisiae*) cells induces the phase transition of amylopectin, even though it has not been subjected to the action of DBEs. Overexpression of ESV1 in an ISA1-deficient background, while inducing the production of minute amounts of insoluble glucans, is not as effective as LESV. However, ESV1 overexpression in a wild-type background induces a high accumulation of insoluble starch. These results have been interpreted to mean that LESV is involved in the phase transition of amylopectin, while ESV1 stabilizes starch granules, protecting them from premature digestion (Liu et al. 2023).

Protein sequence analysis of ESV1 and LESV failed to identify already known catalytic or functional domains (Feike et al. 2016). ESV1 and LESV share a conserved domain of about 240 amino acids located at their C-termini. Their N-terminal regions are of different lengths (130 amino acids in ESV1 and 304 in LESV including the plastid localization signal peptide) and do not share sequence homology to each other. The C-terminal domains contain numerous tryptophan and aromatic amino acid residues organized in conserved repeated motifs also containing acidic amino acids, i.e. aspartic acid and glutamic acid residues. The presence of these repeated motifs could constitute binding sites for numerous glucans or mediate interaction with long glucans such as starch components (Feike et al. 2016). Both proteins were recently investigated by a combination of structural and functional studies (Liu et al. 2023). The structures of ESV1 and LESV were modeled using AlphaFold2 (Jumper et al. 2021), complemented by biophysical approaches (Liu et al. 2023). The results showed a unique and common fold for the conserved C-terminal domain. The tryptophan-rich regions of both ESV1 and LESV proteins have been predicted with high confidence, to fold into an extended planar β -sheet. Localizing the conserved motifs of aromatic and acidic residues within these predicted structures revealed that they align into linear stripes regularly spaced and running across both sides of the β -sheet, perpendicular to the β -strands. The aromatic stripes are about 70 Å long, and distance between them is about 14 Å, which corresponds well to the lengths and spacing, respectively, of the double helices of amylopectin in the crystalline phase of starch granules (Buleon et al. 1998). It has been proposed that this domain constitutes a previously uncharacterized carbohydrate binding surface capable of binding at least 2 double helices of amylopectin molecule on each side of the β -sheet. Synthetic biology approaches in yeast and in vivo experiments in Arabidopsis provide direct evidence that LESV is directly involved in promoting the phase transition of amylopectin. To do that, the described carbohydrate binding surface would allow LESV to bind several double helices of amylopectin, thereby promoting their organization and transition from a soluble to a crystalline phase. This domain would allow ESV1 to maintain the organization of glucans in newly formed granules and to limit their enzymatic degradation during the daytime phase (Liu et al. 2023). Thus, while the 2 proteins share a domain that enables them to bind to ordered double helices of amylopectin, they appear to have different functional roles in the cell. Their N-terminal domains, which are not conserved and whose structure is unknown, could be at the origin of these differences and may modulate polyglucan binding or interactions with other protein partners during starch biosynthesis.

Interestingly, in vitro studies of the affinity of the 2 proteins for starch polyglucans have suggested that ESV1 interacts with amylopectin but not LESV, which binds only amylose (Malinova et al. 2018; Singh et al. 2022). This latter result is inconsistent with the results of Liu et al. (Liu et al. 2023) which show that LESV intervenes in the structuring of amylopectin molecules to form starch granules and should be able to interact with amylopectin. In an attempt to resolve this ambiguity and further elucidate the mechanism of action of these proteins, we studied in the interaction of ESV1 and LESV with α -polyglucans using a combination of biochemical, biophysical, and structural approaches. Our aim was first to determine the affinity of the 2 proteins for the different polyglucans and to identify differences that would allow to clarify their function in the plant. As the N-terminal domains of ESV1 and LESV may be involved in the function of each of the 2 proteins, we also carried out a structural and biophysical study to analyze their potential involvement in the interaction with polyglucans. Our results show that ESV1 and LESV both interact with amylopectin but not with amylose. They further show that this interaction involves conformational changes in the N-terminal domain of LESV. Thus, this domain is also likely involved in the specific recognition of amylopectin by LESV during starch biosynthesis, supporting the idea that it acts upstream of ESV1 in the pathway.

Results

The N-terminal domain of LESV contains helices whose folding and position depend on protein environment

C-terminal domain shared by ESV1 and LESV has an original structure that exhibits the characteristics required for the binding of long polyglucan chains (Liu et al. 2023). However, both proteins have an N-terminal domain, longer in LESV, whose function is not yet defined and for which AlphaFold2 does not have a reliable structure prediction.

To go beyond the AlphaFold2 models available for LESV and ESV1, we recalculated a set of 5 AlphaFold2 models for both proteins. Specifically, we calculated in-house the templates for the constructs used in the CD and SAXS experiments (Liu et al. 2023). The predicted regions with a high degree of confidence (pLDDT > 90) for ESV1 (amino acids from 142 to 395) and LESV (amino acids from 318 to 573) remain identical in the 5 models (Liu et al. 2023) (Fig. 1A) which suggests a high stability for this domain. In these models, we were interested in the predicted structures of the less conserved domains. The regions close to the C-terminal β -sheet of ESV1 (46 amino acid residues at the N-terminus and the polyproline region at the C-terminus), which are poorly conserved between species, are predicted to be disordered. In the C-terminal region of the β -sheet, located on one of its faces (Face A; Fig. 2), a short α -helical region (residues 378 to 395) is predicted with a high degree of confidence for its structure and position relative to the β -sheet. This helix is also present in LESV (residues 555 to 578) with the same degree of confidence.

The structure of the N-terminal region of LESV is predicted with low to very low confidence. However, it has been shown that 3 helical regions located in an island of conservation are predicted with pLDDT > 70 (Liu et al. 2023). Only one of these forms a long helix (residues 245 to 273) whose position relative to the β -sheet is predicted with high confidence on the Face A of the β -sheet (Fig. 1A). It is predicted with the same confidence in all models generated by AlphaFold2, but its position may vary from one model to another, suggesting that it may be modified according to the protein's environment. For other helices, although predicted with pLDDT > 70, their size and positions are not equivalent between models. This result indicates that the N-terminal domain of LESV is probably mainly disordered and susceptible to induced helix folding depending upon the conditions.

We analyzed the structure of the Face A of the β -sheet of ESV1 in the region equivalent to that occupied by helix 245 to 273 on LESV (Fig. 2). AlphaFold2 does not predict a long helix at this location but a loop (amino acids 109 to 138) with a pLDDT > 70 and an expected positional error < 3 Å (Fig. 2). The helix in LESV and loop in ESV1 are stabilized by numerous interactions with the amino acids of the β -sheet in both proteins, including some of the conserved aromatic and acidic residues that cover half the height of the sheet. These 2 structures are located on the same side of the β -sheet as the short, C-terminal helices conserved in both proteins. Their presence in this configuration is not compatible with the binding of the amylopectin double helices, potentially resulting in a polarity in the glucan-binding domain: one side is accessible, and the other is not.

The N-terminal domain of LESV is partially disordered and folds close to the tryptophan-rich domain

The 2 proteins have different functions within the plant, and we hypothesized that this difference might be related to their differing N-terminal domains. These domains are predicted to have rather dynamic structures—a property incompatible with structural analysis by X-ray crystallography. Therefore, we investigated their structures using a combination of more appropriate biophysical approaches such as circular dichroism and SAXS.

Small angle X-ray scattering (SAXS) is an X-ray scattering approach for proteins in solution. While this approach does not provide high-resolution structure, it allows the analysis



Figure 1. Structure of LESV of A. *thaliana*. The structure is represented in cartoon. **A**) Superposition of the 5 molecular models of LESV calculated with AlphaFold2. Only regions with pLDDT > 70 are shown except for the first model for which unfolded parts are visible. Regions common to all 5 models with pLDDT > 90 are colored dark gray. The helices of the 5 models are colored in light gray, salmon, cyan, yellow, and green. The common helix with ESV1 is on the left, and the long helix specific for LESV is on the right. **B**) Ab initio model of the N-terminal domain of LESV computed from SAXS data using the software BUNCH. The conserved C-terminal domain is represented as cartoon and the N-terminal domain model is represented as spheres (one sphere by amino acid residue). **C**) Superposition of SAXS experimental data obtained for LESV (cyan) and calculated curve from BUNCH model (black and smoth line) with $\chi^2 = 3.8$ Å.

of the molecular envelope and the position of protein domains in relation to each other. If the structure of 1 domain is known, it can be used to position and model ab initio the structure of the missing part, which can be crucial in studying the function of a dynamically structured protein. We used SAXS to analyze the structure and position of the N-terminal domain of LESV, as that of ESV1 is extremely short.

In order to visualize and localize the N-terminal domain of LESV, we performed an *ab initio* modeling based on the structure of the high-resolution C-terminal domain model given by AlphaFold2 and the SAXS data (Liu et al. 2023). The result obtained for LESV is shown in Fig. 1B. The obtained model, which fits the SAXS data with a high degree of confidence ($\chi^2 = 3.8$) (Fig. 1C), shows that the end of the C- and N-terminal domains (between 50 and 80 residues) is rather disordered and emerges from the overall structure, while the rest of the domain is organized around or close to the β -sheet. The presence of this domain around, or at least close to, the β -sheet is likely to affect glucan binding and could contribute to the differences in function between the 2 proteins described in Liu et al. (2023).

ESV1 and LESV interact differently with α -1,4-linked glucose polymers

The structural polydispersity of amylose and amylopectin solutions, as well as their high viscosity, precluded their use in conventional structural biology approaches (SAXS or X-ray crystallography) or SPR. To clarify the specificities of ESV1 and LESV and their interaction with starch glucans, we performed electromobility shift assay (EMSA) experiments which are better suited to the biochemical properties of glucans. EMSA is a rapid and sensitive method to detect protein-glucan interactions. It is based on the observation that the electrophoretic mobility of a protein can be retarded in polyacrylamide gels containing increasing concentrations of a ligand, leading to a shift in the position of the protein band. In case of specific affinity, the intensity of this shift will be proportional to the concentration of glucan in the gel. First, we followed the influence of increasing concentrations of amylose or amylopectin on the electrophoretic mobility of LESV and ESV1 (Fig. 3). On native polyacrylamide gels containing 0.1% and 0.3% (w/v) amylopectin, both ESV1 and LESV show a large reduction in mobility which increased with amylopectin concentration (Fig. 3A), demonstrating a strong affinity of the 2 proteins for this polysaccharide.

In contrast, no electrophoretic mobility differences were found for LESV and ESV1 in native gels containing amylose in the same concentration range, suggesting that LESV and ESV1 have no affinity for amylose under the tested conditions (Fig. 3B). Amylose contains longer chains than amylopectin and much fewer branching points. Next, we decided to test the affinity of LESV and ESV1 for glycogen (a highly



Figure 2. Structure of C-terminal domains of LESV and ESV1. **A)** Conserved structural motifs on Face A of LESV and **B)** ESV1. The structures are represented as cartoon, the common β -sheet is colored in magenta, the common C-terminal helix is on the left of the figures (left panel) or in the foreground (right panel) colored in cyan, and the long helix of LESV (**A**) and the long loop of ESV1 (**B**) are colored in yellow. The right panel is another view of the left panel after a rotation of 90° along the *y* axis.

branched α -linked glucans) (Fig. 3C). Two different final concentrations of glycogen (0.25% and 1% w/v) were added to 8% (w/v) acrylamide native gels. A shift to lower mobility of the LESV protein band was observed in the gel containing 0.25% (w/v) glycogen compared with the reference protein band. This shift is accentuated when the glycogen concentration is increased, suggesting here again that LESV has specific affinity for this branched glucan. Interestingly, ESV1 did not show any change in its electrophoretic mobility in the presence of glycogen, eliminating a potential affinity of ESV1 for glycogen.

Binding of amylopectin to LESV causes α -helices to appear in the N-terminal domain of the protein

In order to better characterize the mode of interaction of ESV1 and LESV with glucans and the role of N-terminal domains and to highlight conformational changes of the proteins' structure during complex formation, we performed an Synchrotron radiation circular dichroism (SR-CD) study. CD is the method of choice for analyzing the structure of a protein by visualizing its content in secondary structural elements. It also allows the study of interactions between proteins and their ligands, especially when the latter leads to protein structural changes. SR-CD extends the limits of typical CD spectroscopy by providing an extended spectral range, improving signal-to-noise ratio and enabling faster data acquisition (Hussain et al. 2012; Hussain et al. 2018). Quantitative analysis of CD spectra also makes it possible to predict the secondary structure content of a protein.

We compared the spectra obtained for the proteins alone and in the presence of amylopectin and amylose as the latter does not interact with the proteins. The CD spectra of both proteins alone have been described in Liu et al. (2023) and show that they are structured albeit with differences in the secondary structure composition (Figs. 4 and 5). For ESV1, the pattern of CD spectra corresponds to a folded protein with a strong positive band at $\lambda = 196$ nm and only 1 negative band at $\lambda = 220$ nm, which are characteristic of proteins containing mainly β -strands/sheets. For LESV, the pattern of the spectrum reveals a global folding of β -strands and α -helices. Indeed, the LESV CD spectrum shows a strong maximum at $\lambda = 192$ nm and a minimum at $\lambda = 216$ nm, which are the signature of the presence of β -structures, but unlike ESV1, the spectrum also shows 2 shoulders at $\lambda = 210$ and $\lambda = 222$ nm, which is evidence for the additional presence of α -helices (Liu et al. 2023).

To analyze the structural effect of interactions between the ESV1 and LESV proteins and amylopectin, solutions of each



Figure 3. EMSA gels analyzing the interaction between ESV1 or LESV and starch glucans. **A)** Interaction with amylopectin, **B)** interaction with amylose, and **C)** interaction with glycogen. Blue, red, and yellow arrows indicate the bands corresponding to the reference protein, ESV1 and LESV, respectively. The migration shift is indicated by a black arrow.



Figure 4. SR-CD spectra for ESV1 alone (gray) and in the presence of amylopectin (red) or amylose (blue). The composition in secondary structural elements evaluated by BestSel is in inset. NRMSD is the normalized root mean square deviation.

protein were mixed with 1% (w/v) amylose or 1% (w/v) amylopectin solutions prior to measurement. For each spectrum, the composition of the secondary structure elements was determined using BestSel (Micsonai et al. 2015). The values obtained were compared with those obtained for the proteins alone.

Figure 4 shows the superposition of the spectra of ESV1 alone and added to amylopectin or amylose solutions.

The analysis shows broadly equivalent spectra for ESV1 in the presence of amylose or amylopectin, with a degree of structuring that appears to be slightly lower for the protein alone. Analysis of the composition of secondary structural elements has been performed using BestSel (Fig. 4, inset). BestSel is a tool that allows the secondary structure determination and fold recognition from protein circular dichroism spectra. It indicates an equivalent composition for



Figure 5. SR-CD spectra for LESV alone (gray) and in the presence of amylopectin (red) or amylose (blue). The composition in secondary structural elements evaluated by BestSel is in inset. NRMSD is the normalized root mean square deviation.

ESV1 alone or in the presence of amylose and a slight reduction or modification of some strands and α -helices when the protein is in the presence of amylopectin. This result indicates that ESV1 interacts with amylopectin without undergoing conformational changes in the protein. This result demonstrates that the C-terminal domain, which constitutes the vast majority of ESV1, does not undergo a conformational change following interaction with amylopectin.

The same analysis was performed for LESV (Fig. 5). The spectrum obtained for the LESV/amylopectin mixture shows peaks of much higher magnitude than those obtained for the protein alone or in the presence of amylose. The positive peak at $\lambda = 192$ nm has a magnitude 60% higher than that of the protein alone, indicating a higher structuring of the protein. At $\lambda = 208$ and $\lambda = 215$ nm, the molar ellipticity values indicating the presence of α and β structures are 50% lower than those observed for the protein alone or in the presence of amylose. More interestingly, the molar ellipticity at $\lambda = 222$ nm, indicating the presence of α -helices, is much lower (70%) than that observed for the protein alone or in the presence of amylose. This result shows that the binding of amylopectin induces a conformational change of LESV notably through the formation of additional α -helices upon binding of amylopectin.

To identify and quantify the conformational changes undergone by LESV in the presence of amylose and amylopectin, the composition of the secondary structural elements was analyzed using BestSel (Micsonai et al. 2015). The results (Fig. 5, inset) confirm the analysis of the CD spectra: the composition of the β -strands and turns is equivalent whether LESV is alone or in the presence of amylose or amylopectin, likely indicating that the structure of the β -domain is not modified by the presence of polyglucans. The high conservation observed in the C-terminal domain of both LESV and ESV1 supports the hypothesis that the lack of conformational change in the ESV1 C-terminal domain upon amylopectin binding confirms this hypothesis. A higher number of α -helices is observed when LESV is in the presence of amylopectin. This confirms that LESV interaction with amylopectin induces the formation of α -helices, presumably within the N-terminal domain, with the structure of the C-terminal domain remaining unchanged.

The spectrum obtained for the LESV/amylose mixture had a broadly similar appearance to that of LESV alone, with peaks of slightly lower magnitude confirming EMSA experiments showing no affinity for this glucan. LESV in the presence of amylose seems to have a slightly lower number of α -helices and strands than the protein alone. This result suggests that although LESV does not interact with amylose, the presence of high amounts of amylose may slightly modify the behavior of the protein in solution.

Amylopectin binding affects the melting temperature (TM) of LESV but not that of ESV1

Enhanced detection of ligand binding can be achieved through thermal denaturation studies monitored by SR-CD. This method is more sensitive than simple spectral differences as it can detect interactions that do not induce structural modifications of the proteins. The experiment involves measuring the CD spectrum of proteins at different temperatures alone or in the presence of ligand. Increasing the temperature causes progressive denaturation of the protein and therefore a change in the molar elipticity. These changes can be used to monitor denaturation at a given wavelength and estimate the mixing temperature (TM). As the presence of a ligand generally tends to stabilize the protein, its TM will be higher in the presence of the ligand.

The CD signal variation was measured as a function of the temperature for both proteins, alone and in the presence of amylopectin, under the same conditions as for the spectra at



Figure 6. Thermal denaturation of LESV and ESV1 followed by the variation of the SR-CD molar ellipticity in function of the temperature. Plots represent consecutive scans on the protein collected at a set of temperature between 20 and 90 °C colored in a gradient from dark blue (lowest temperature) to light blue (highest temperature) for **A**) LESV alone, **B**) LESV with amylopectin, **C**) ESV1 alone, and **D**) ESV1 with amylopectin. From these spectra, molar ellipticity at wavelength indicated by an arrow on the scans and molar ellipticity in function of the temperature have been used to monitor **E**) thermal denaturation of LESV followed at $\lambda = 195$ nm and **F**) thermal denaturation of ESV1 followed at $\lambda = 190$ nm. Curves corresponding to proteins alone or combined with amylopectin are colored in gray and red, respectively.

constant temperature. Figure 6, A to D, present the different spectra obtained for LESV and ESV1 during the temperature gradient. To assess the impact of the presence of amylopectin on LESV and ESV1 stability, we measured the TM of the mixtures by monitoring the molar ellipticity evolution as a function of temperature at $\lambda = 195$ and $\lambda = 190$ nm, respectively. The curves obtained have been normalized and are presented in Fig. 6, E and F. The experiment shows that the denaturation of LESV is slowed down in the presence of amylopectin as the TM increases dramatically from 55° to 65°. Thus, the presence of amylopectin stabilizes LESV, attesting to a strong interaction. The denaturation curves for ESV1 show the same pattern and can be overlaid with an inferred TM of about 50°C for ESV1. Therefore, in contrast to LESV,

the binding of amylopectin does not affect the thermostability of ESV1.

ESV1 and LESV accumulate on the entire surface of starch granules

Having demonstrated the interaction between ESV1 and LESV in solution, we investigated whether the proteins are also able to bind amylopectin in its insoluble form. To do that, the binding of ESV1 and LESV to starch granules was monitored by UV fluorescence microscopy. This approach visualizes proteins *via* the fluorescence of their aromatic residues without adding any external probe and has already been used to visualize proteins on starch granules



Figure 7. Transmitted light and fluorescence imaging of maize waxy starch granules in the presence of ESV1 (top panel), LESV (medium panel), and BSA (low panel) as negative control. **A)** Visible light imaging, **B)** fluorescence images of ESV1 or LESV absorption on starch granules, and **C)** combination of visible light and fluorescence images. (scale bar for all images: 50 μm).

(Tawil et al. 2011). To visualize only ESV1 and LESV, we used starch granules from waxy maize that has no GBSS protein present in the starch granules (GBSS is the major granulebound protein). Thus, the fluorescence of the granules could be distinguished from the fluorescence of the protein being studied (see Materials and methods). The measurement was carried out simultaneously in visible light, to visualize starch granules, and with excitation at $\lambda = 310$ nm, which allows the tryptophan residues to be excited, to visualize proteins (Fig. 7). The emission spectrum was obtained using a filter to select a wavelength range between 329 and 351 nm which is specific to tryptophan residues. Two controls were carried out, the first with starch granules alone to verify the absence of fluorescence and the second with starch granules incubated with BSA to verify the absence of unspecific protein binding. When starch granules were incubated with ESV1 or LESV proteins, the tryptophan fluorescence images revealed a distinct halo over the surface of the starch granules, demonstrating the affinity of the proteins for amylopectin in its insoluble, semicrystalline form. However, the intensity of the fluorescence seemed to be greater in the case of ESV1 particularly on larger granules. This could be interpreted as ESV1 having a greater affinity than LESV for amylopectin in its insoluble form.

The C-terminal tryptophan-rich domain of ESV1 and LESV can bind 2 double helices of amylopectin (at least) on 1 face of the β -sheet

To gain more insight into the binding of ESV1 and LESV to insoluble amylopectin, we simulated the interaction between the C-terminal domain of the proteins and 2 double helices of amylopectin. To do this, we used the model of LESV containing the β -sheet and the conserved and well-positioned



Figure 8. Molecular model of the complex between C-terminal domain of LESV and amylopectin double helices. Protein chain is represented in cartoon and colored in blue. Romatic residues are colored in magenta, and their side chains are represented by sticks. Amylopectin double helices are represented by sticks and colored by atom types.

 α -helix described in Liu et al. (2023) and shown in Fig. 1. Since in both proteins one side of the C-terminal β -sheet may be partially obscured (by the long helix in LESV and the long loop in ESV1; Fig. 2), we only studied the interaction of amylopectin on the accessible side of the β -sheet. For the docking simulation, the parts of the N-terminal domain other than the conserved, well-positioned α -helix were omitted.

We first performed a docking calculation with 1 molecule of protein and 1 double helix of amylopectin centered on 1 aromatic stripe of LESV on the accessible face of the β -sheet. We obtained a good solution in which the amylopectin double helix binds the β -sheet, aligning well with the aromatic stripe. We repeated the same approach with the protein binding a second amylopectin double helix as the target centered on the second aromatic stripe. We again obtained a solution shown in Fig. 8. On this structure, 2 double helices of amylopectin bind along the aromatic stripes and lie parallel to each other separated by 10 Å (between the axis of the double helices). This arrangement of the double helices in relation to each other corresponds to the arrangement of the amylopectin molecules described for starch (Imberty et al. 1988). This result suggests that the conserved C-terminal domain conserved of ESV1 and LESV is perfectly suited for the interaction of the proteins with the insoluble form of native amylopectin or for helping that structure to form.

The amylopectin molecules interact with the protein domain through numerous interactions typical of protein–sugar interactions, as predicted from the analysis of the primary sequence of the β -sheet domain and the distribution of conserved amino acids in the AlphaFold2 model (Fig. 8). The glucose rings interact by hydrophobic stacking with the aromatic rings of the β -domain all along the double helices. The acidic residues conserved in LESV and ESV1 also play

an important role in the interaction by participating in hydrogen bonding with the hydroxyl groups of the glucose residues on the sides of the double helices. On Face A, described above, there is an analogous organization in stripes of aromatic and acidic residues, suggesting that double helices could also bind on this face, possibly after structural reorganization of the conserved helix (LESV) or loop (ESV1), allowing both sides of the proteins to interact with amylopectin molecules.

Discussion

ESV1 and LESV bind specifically to amylopectin

In this work, we investigated the interaction specificities of ESV1 and LESV with the different components of starch. The properties of amylopectin and amylose macromolecules, which are complex, nonhomogeneous, and often dense glucans in solution, limited the approaches that could be used. Previous work on the affinity of ESV1 and LESV for starch components (Malinova et al. 2018; Singh et al. 2022) focused on the interaction between ESV1 and LESV and starch glucans in insoluble form and/or on different mutant starch granules. That work proposed that ESV1 and LESV interact with starch granules, each having specific affinity for amylopectin or amylose respectively, with affinity being independent of the protein/glucan ratio. Those results are, however, not fully consistent with the recent characterization of LESV and ESV1 (Liu et al. 2023) nor with the results presented here. First, we chose an EMSA approach, which allowed us to analyze the behavior of both proteins in relation to each of the polyglucans. The migration profile of the 2 proteins shows that both proteins have a strong affinity for amylopectin and that this affinity increases with the amount of this glucan, attesting to their specificity. Conversely, in the presence of amylose, we did not observe any migration

retardation associated with amylose. Second, structural analysis of ESV1 and LESV in the presence of amylose and amylopectin by SR-CD shows that only the presence of amylopectin induces conformational changes in LESV upon interaction, leading to the structuring of disordered regions of the N-terminal domain of the protein into α -helices. No conformational changes were observed in ESV1, but this can be explained by the fact that the protein has a reduced N-terminal domain and consists almost entirely of the highly structured C-terminal changes.

As the experiments we carried out were with amylopectin molecules in its solubilized form, we wanted to verify that ESV1 and LESV were able to interact with the crystallized form. The results we obtained in fluorescence microscopy with maize *waxy* starch granules, which contain no amylose, showed that ESV1 and LESV interact directly with amylopectin starch granules. In fact, the 2 proteins, identified by their inherent fluorescence, accumulated on the entire surface of the starch granules. This result confirms those obtained in solution and is consistent with the role described in our recent work (Liu et al. 2023).

LESV is able to bind to amylopectin during its biosynthesis

In this work, we further showed that the 2 proteins behave differently toward glycogen as only LESV was able to interact with it. This finding is very interesting on several levels. First, it shows that the difference in affinity of ESV1 for amylopectin and amylose does not seem to be related to the presence of branch points but rather to the 3D structure of the glucan. Secondly, an analogy has been drawn between the structure of glycogen and that of the precursor molecules of amylopectin during its biosynthesis, before the action of isoamylases that remove the excess branching points (Ball et al. 1996). This result strengthens the proposed function of LESV in our previous work (Liu et al. 2023). On the one hand, LESV could function concomitantly with isoamylases during the biosynthesis of amylopectin-supporting the phase transition of double helices (from soluble to semicrystalline form), as the branching pattern is optimized. However, on the other hand, we showed that LESV can also promote the phase transition of nascent amylopectin even if it is not debranched by isoamylase and would otherwise remain soluble as phytoglycogen (Liu et al. 2023). Our results are also consistent with the idea that ESV1 functions downstream of LESV, stabilizing newly formed starch granules. Thus, while ESV1 has an affinity for the amylopectin after the isoamylases have optimized its structure for crystallization, it has a low affinity for glycogen, which will not undergo phase transition of its own accord.

LESV and ESV1 share a common domain whose structure was modeled and described as being particularly compatible with the binding of amylopectin double helices. The fact that ESV1, which is predominantly composed of this domain, does not interact with glycogen suggests that the interaction with the nascent amylopectin is mediated or assisted by the N-terminal domain of LESV, giving it additional specificity and allowing it to accommodate different glucans than ESV1. Indeed, we previously noted that the N-terminus of LESV also has highly conserved aromatic amino acids in parts of the protein that are predicted either to be in α -helical structure or to be unstructured (Liu et al. 2023). It will be interesting to examine the roles of these residues in future studies.

The C-terminal domain is able to bind at least 2 double helices of amylopectin

The models obtained for ESV1 and LESV via AlphaFold2 provided a very reliable structure for the C-terminal tryptophanrich domain, which is conserved between the 2 molecules. The other parts of the molecule were predicted without much reliability (Liu et al. 2023). The C-terminal domain folded into an original structure, forming a rather large oval (about 40 Å wide and 70 Å long) antiparallel twisted β -sheet. On this β -sheet, the aromatic and acidic residues, organized in repeated sequences identified during the analysis of the protein sequences, form parallel lines equidistant from each other and parallel to the axis of the β -sheet. The side chains of these amino acids point alternately to both sides of the β -sheet. However, 1 face of the β -sheet may be "occupied" by a long α -helix in LESV and a long loop in ESV1. Therefore, we computed models where only the "free" side interacted with amylopectin double helices. Even though we observed protein conformational changes in the presence of amylopectin, based on our current data, we do not know if the protein parts on the occupied face can move to expose the lines of aromatic and acidic residue for additional amylopectin binding. If that does occur, ESV1 and LESV could bind amylopectin on both sides of their β -sheet, resulting in sandwich-like alignments of proteins and amylopectin.

We were able to demonstrate experimentally that LESV and ESV1 can bind amylopectin in both its soluble form and its organized form within starch granules. Furthermore, our modeling work shows that this interaction likely occurs through the shared β -sheet domain via both the aromatic and acidic residues. Moreover, we suggested that the unique structure of this domain enables it to bind at least 2 parallel double helices of amylopectin in a parallel arrangement similar to that found in the crystalline regions of starch granules, thereby supporting its proposed function in the organization and maintenance of starch granules in plants.

The N-terminal domain allows regulation of LESV specificity toward starch components

The structure of the N-terminal domain of LESV is not known. However, we have been able to demonstrate that it consists mainly of disordered regions and α -helices that may be organized close to the C-terminal domain. We have also shown that interaction with amylopectin induces the

formation of α -helices presumably from the disordered regions without affecting the β -sheet structure. In ESV1, this domain N-terminal domain is shorter, and poorly conserved between species, but still predicted to be unstructured, as is the polyproline tail at the C-terminus of the Arabidopsis protein. It is likely that the presence of these regions prevented us from obtaining protein crystals. Since amylopectin structure and size is not monodisperse, it cannot be used to stabilize the proteins, and we are currently searching for analogs that can facilitate the crystallization of both proteins. While analyzing the structure of the C-terminal domains has allowed us to describe the interaction mode of the 2 proteins with amylopectin, analyzing the entire structures, particularly for LESV, would help us to elucidate the function of the N-terminal domain, potentially revealing the mechanism of LESV action in promoting amylopectin crystallization and explaining the difference between it and ESV1. Considering that several helices are present in the LESV N-terminus and taking into account its involvement in amylopectin biosynthesis, it is possible that LESV may also interact with other proteins, as already described for SSs and BEs (Ahmed et al. 2015; Crofts et al. 2015).

In conclusion, this work improves our understanding of the molecular mechanisms of ESV1 and LESV function in Arabidopsis. We have shown that the conserved C-terminal domain of both ESV1 and LESV is particularly well suited to bind amylopectin double helices as they are organized in starch granules. We further propose that LESV is able to interact with nascent amylopectin molecules during its biosynthesis and that its involvement in the phase transition probably occurs before the biosynthesis process is complete. Our data are consistent with the idea that ESV1 would intervene later to stabilize the granules to prevent early degradation by hydrolytic activities. Further research is needed to describe the precise molecular mechanism of ESV1 and LESV function in plants. Resolution of the atomic structures of the entire LESV, particularly the organization of its N-terminal domain in relation to its glucan-binding domain and its interaction with starch glucans, will allow considerable progress to be made in this area. Finally, studying the function of these proteins in other plants-specifically their involvement in storage starch biosynthesis-will be important to assess their candidacy as targets for starch crop improvement through biotechnological approaches.

Materials and methods

Cloning, expression, and purification of proteins

LESV and ESV1 from Arabidopsis (*Arabidopsis thaliana*) were cloned, expressed in *Escherichia coli* as recombinant proteins lacking their N-terminal transit peptides, and purified as described previously. The purification batches of proteins used in this work are the same than those used in previous work (see Supplementary Fig. S2 in Liu et al. (2023)). Purification was performed by a first step of Immobilized Metal Affinity Chromatography followed by a second purification step through size exclusion chromatography using a HiLoad 16/ 60 Superdex 200 (Cytiva) column preequilibrated with 50 mm Tris-HCl, pH 8, 150 mm NaCl, 10% (w/v) glycerol, and 2 mm DTT for LESV or a dialysis step against 50 mm Tris-HCl, pH 7.5, 100 mm NaCl, 10% (v/v) glycerol, and 2 mm DTT for ESV1. The monodispersity of the obtained protein solution was assessed by dynamic light scattering (DLS) using a zetasizer pro (Malvern Panalytical). For structural study, protein samples were concentrated using Vivaspin centrifugal concentrator with a 10 kDa cutoff (Sartorius). Protein concentrations were determined using a NanoDrop Spectrophotometer (ND1000) from Thermo Scientific.

Glucan solution preparation

For EMSA experiments, amylose 1% (w/v), amylopectin 1% (w/v), and glycogen 5% (w/v) stock solutions used for these experiments were prepared as follows. 0.1 g of amylose (from potato [S. tuberosum]; Sigma] was dissolved in 1 mL of 2 M sodium hydroxide (NaOH) and vortexed to ensure complete solubilization. After addition of 2 mL of deionized water, the solution was neutralized by 1 mL of 2 M HCl. The final volume was then adjusted to 10 mL with water and the solution was heated 5 min at 50 °C and vortexed until complete dissolution. Amylopectin (0.1 g, from potato, Sigma) was resuspended in 10 mL of distilled water and then subjected to autoclaving to produce a homogeneous solution. Glycogen powder (0.5 g, from oyster [Ostrea edulis]; Sigma) was dissolved by vortexing it in distilled water (final volume 10 mL) until full homogenization. For CD experiments, glucan stock solutions were prepared with the same protocol albeit replacing water with protein buffer.

EMSAs

One microgram of ESV1, LESV, and a reference protein with no affinity for glucans (UniProt Q7W019 produced in the lab (Herrou et al. 2007)) were loaded on 8% (w/v) polyacrylamide gels containing increasing concentrations of glucans (from 0.1% to 0.3% [w/v] for amylopectin and amylose and 0.1% to 1% [w/v] for glycogen) and submitted to electrophoresis in native conditions at 4 °C at 15 V/cm for 2 h in 25 mM Tris-HCL, pH 8.3, 192 mM glycine migration buffer. Gels were stained with InstantBlue (Expedeon). The affinity of LESV and ESV1 for the different glucans in the gels was estimated by the migration shift of these proteins compared with the stable migration of the reference protein, either loaded alone or mixed with ESV1 or LESV.

SR-CD

SR-CD spectra were measured at the DISCO beamline of the SOLEIL Synchrotron (Gif-sur-Yvette, France). Five microliters of ESV1 protein at 6.1 mg/mL and 2 μ L of LESV protein at 13.4 mg/mL were deposited between 2 CaF₂ coverslips with a pathlength of 20 and 10 μ m, respectively (Refregiers et al. 2012). The beam size of 4 × 4 mm and the photon-flux per nm step of 2 × 10¹⁰ photons/s in the spectral band from 270 to 170 nm prevented radiation-induced damage

(Miles et al. 2008). CD spectra were acquired using IGOR software (WaveMetrics). Protein and buffer spectra were collected consecutively and are the mean of 3 accumulations. The buffer baseline was recorded sequentially and subtracted from the spectra before taking into account the protein concentration. Before measurements, the molar elliptical extinction coefficient of ammonium d-10-camphorsulfonate ammonium (CSA) has been measured on the beamline and used as standard for calibration of all data measurements (Miles et al. 2004). Data processing was conducted using CDToolX software (Miles and Wallace 2018). The influence of the different glucans on the structure of LESV and ESV1 was studied by incubating the protein/glucan mixtures for 2 h and measuring the spectra under the same conditions as for the native proteins. The mixtures were made with 4:1 mix of protein solution and glucan solution (1% [w/v] a mylose or a mylopectin solutions). Five microliters of ESV1/glucans mixtures and 2 μ L of LESV/glucan mixtures were deposited between 2 CaF₂ coverslips with a pathlength of 20 and 10 μ m, respectively (Refregiers et al. 2012). Spectra containing 80% v/v of protein buffer and 20% glucan solutions were subtracted from the protein/glucan spectra before CSA calibration. Temperature scans were realized to check the protein stabilization by glucan interaction. CD spectra were collected from 20 to 30°C to 90°C and processed as described above with 3 to 5°C temperature increases and 3 min of equilibration time. The secondary structure element content of each protein alone or in the presence of glucans was estimated using BestSel (Micsonai et al. 2015).

Molecular modeling

Protein structures of LESV and ESV1 were modeled using AlphaFold2 (Jumper et al. 2021). For each protein, 5 different models were computed and ranked by global predicted local distance difference test (pLDDT). The 5 molecular models generated were superimposed and used to evaluate the possible position of dynamic regions. Molecular models with best pLDDT values were used for figures and further molecular docking studies.

SAXS

Protein sample solutions were centrifuged for 10 min at 10,000 × g prior to X-ray analysis as a precaution to remove any insoluble aggregates. SAXS experiments were conducted on the SWING beamline at Synchrotron SOLEIL ($\lambda = 1.033$ Å). All solutions were mixed in a fixed-temperature (15°C) quartz capillary. The monodisperse sample solutions of proteins were injected onto a size exclusion column (David and Perez 2009) (SEC-3, 150 Å; Agilent) using an Agilent HPLC system and eluted into the capillary cell at a flow rate of 0.3 mL/min. Then, 50 μ L of protein samples were injected for SAXS measurements. A total of 180 frames were collected during the first minutes of the elution and were averaged to account for buffer scattering and subtracted from selected frames corresponding to the main protein elution peak. Data reduction to absolute units, frame averaging,

and subtraction were done using FOXTROT (David and Perez 2009). Data processing, analysis, and modeling steps were carried out using programs of the ATSAS suite (Franke et al. 2017). BUNCH (Petoukhov and Svergun 2005) was used to model the missing parts of the proteins that were not assigned by AlphaFold2 (Jumper et al. 2021). The fit of the model obtained with BUNCH to the experimental SAXS data is estimated by superimposing the SAXS curve derived from the model on the measured SAXS curve. A low χ^2 value indicates a good superposition of the curves and therefore that the model is compatible with the experimental data.

Docking

AlphaFold2 model structures of the conserved C-terminal domain of ESV1 and LESV were used as targets in order to model the binding position of the model of a double helix of amylopectin obtained from Polysac3DB (CERMAV; https://polysac3db.cermav.cnrs.fr). A generic algorithm was used for the search step within a sphere of 10 Å centered on the tryptophan stripes of the conserved β -sheet domain. The scoring function was based on the ChemPLP forcefield, as used by GOLD (Jones et al. 1997). All the parameters were kept by default. A subsequent energy minimization was performed on the best model using the Amber forcefield. All figures representing molecular structures of proteins and ligands were generated using Pymol (The PyMOL Molecular Graphics System, version 1.8.0.0 Schrödinger, LLC).

3D imaging

For this experiment, we used waxy maize starch granules (Roquette, France) obtained from plants lacking granulebound starch synthase (GBSS; (Tsai 1974)), the enzyme that catalyzes the biosynthesis of amylose and whose presence in the granules is responsible of their autofluorescence. Starch granules were washed with water, acetone, and ethanol in order to remove any phenol contaminants coming from storage in plasticware (Tawil et al. 2011). ESV1 and LESV binding on starch granules was assessed by both visible and UV microscopy with the same protocol described in Jamme et al. (2014). Excitation was set-up at $\lambda = 280$ nm with an emission filter at 329 to 351 nm (FF01-340/22, Semrock) to visualize the tryptophan emission (at 345 nm) of bound proteins. The acquisition time was 10 s for each emission fluorescence image and 0.2 s for visible images. For 3D purpose, Z slices (along the optical axis) were recorded with a step size of 300 nm over 40 μ m Z range under µManager control (Edelstein et al. 2010). Using imaging analysis software (Huygens, SVI, NL), deconvolution images were calculated by PSF deconvolution treatment. Images were coupled to classical light imaging of the starch granule morphology. Images were analyzed using Fiji (Schindelin et al. 2012). Noise was removed from acquired 3D stacks using a median 1 filter. A substack of "in focus" images were selected and summed. To compensate for field inhomogeneity, a FFT bandbass filter was then applied.

Accession numbers

The Arabidopsis Genome Initiative gene codes for the Arabidopsis genes used in this study are as follows: *ESV1*, *At1g42430*, and *LESV*, *At3g55760*.

Acknowledgments

The authors strongly acknowledge DISCO beamline and the regular access to the SAXS beamline SWING at Synchrotron SOLEIL (St Aubin, France) through the BAG MX-20181002 and MX-20201190 and are grateful for the expert technical support provided by beamline staffs.

Author contributions

C.B. and R.O. conceived and designed the experiment with input from S.C.Z., D.D., and C.d.H. R.O. and C.B. performed the experiments with input from M.B., C.L., and C.S. C.B., R.O., and M.B. collected the Synchrotron data. C.B. and R.O. analyzed the data with the input of M.B. and D.D. C.B. wrote the manuscript. C.B., D.D., C.d.H., S.C.Z., and C.S. revised the manuscript.

Funding

This work has been supported by Centre National de la Recherche Scientifique France (core funding to C.S., D.D., and C.B. and PhD grant for R.O.), Université de Lille France (core funding to C.d.H. and PhD grant to M.B.), and Région Hauts de France (PhD grant for M.B.).

Conflict of interest statement. None declared.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

References

- Ahmed Z, Tetlow IJ, Ahmed R, Morell MK, Emes MJ. Protein-protein interactions among enzymes of starch biosynthesis in high-amylose barley genotypes reveal differential roles of heteromeric enzyme complexes in the synthesis of A and B granules. Plant Sci. 2015:233:95–106. https://doi.org/10.1016/j.plantsci.2014.12.016
- Ball S, Guan HP, James M, Myers A, Keeling P, Mouille G, Buleon A, Colonna P, Preiss J. From glycogen to amylopectin: a model for the biogenesis of the plant starch granule. Cell. 1996:86(3):349–352. https://doi.org/10.1016/S0092-8674(00)80107-5
- Buleon A, Colonna P, Planchot V, Ball S. Starch granules: structure and biosynthesis. Int J Biol Macromol. 1998:23(2):85–112. https:// doi.org/10.1016/S0141-8130(98)00040-3
- Crofts N, Abe N, Oitome NF, Matsushima R, Hayashi M, Tetlow IJ, Emes MJ, Nakamura Y, Fujita N. Amylopectin biosynthetic enzymes from developing rice seed form enzymatically active protein complexes. J Exp Bot. 2015:66(15):4469–4482. https://doi.org/10.1093/ jxb/erv212
- David G, Perez J. Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological

small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. J Appl Crystallogr. 2009:**42**(5):9. https://doi.org/10.1107/S0021889809029288

- Delatte T, Trevisan M, Parker ML, Zeeman SC. Arabidopsis mutants Atisa1 and Atisa2 have identical phenotypes and lack the same multimeric isoamylase, which influences the branch point distribution of amylopectin during starch synthesis. Plant J. 2005:41(6):815–830. https://doi.org/10.1111/j.1365-313X.2005.02348.x
- Edelstein A, Amodaj N, Hoover K, Vale R, Stuurman N. Computer control of microscopes using microManager. Curr Protoc Mol Biol. 2010: Chapter 14:Unit14.20. https://doi.org/10.1002/0471142727.mb1420s92
- Feike D, Seung D, Graf A, Bischof S, Ellick T, Coiro M, Soyk S, Eicke S, Mettler-Altmann T, Lu KJ, et al. The starch granule-associated protein EARLY STARVATION1 is required for the control of starch degradation in Arabidopsis thaliana leaves. Plant Cell. 2016:28(6): 1472–1489. https://doi.org/10.1105/tpc.16.00011
- Franke D, Petoukov MV, Konarev PV, Panijkovich A, Tuukkanen A, Mertens HDT, Kikhney NR, Hajizadeh NR, Franklin JM, Jeffries CM, et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. J Appl Crystallogr. 2017:50(4): 1212–1225. https://doi.org/10.1107/S1600576717007786
- Helle S, Bray F, Verbeke J, Devassine S, Courseaux A, Facon M, Tokarski C, Rolando C, Szydlowski N. Proteome analysis of potato starch reveals the presence of new starch metabolic proteins as well as multiple protease inhibitors. Front Plant Sci. 2018:9:746. https:// doi.org/10.3389/fpls.2018.00746
- Herrou J, Bompard C, Antoine R, Leroy A, Rucktooa P, Hot D, Huvent I, Locht C, Villeret V, Jacob-Dubuisson F. Structure-based mechanism of ligand binding for periplasmic solute-binding protein of the bug family. J Mol Biol. 2007:**373**(4):954–964. https://doi.org/10.1016/j.jmb.2007. 08.006
- Hussain R, Javorti T, Siligardi G, eds. Spectroscopic analysis: synchrotron radiation circular dichroism. Vol. 8. Amsterdam: Elsevier; 2012
- Hussain R, Longo E, Siligardi G. UV-denaturation assay to assess protein photostability and ligand-binding interactions using the high photon flux of diamond B23 beamline for SRCD. Molecules. 2018:23(8):1906. https://doi.org/10.3390/molecules23081906
- Imberty A, Chanzy H, Perez S, Buleon A, Tran V. The double-helical nature of the crystalline part of A-starch. J Mol Biol. 1988:201(2): 365–378. https://doi.org/10.1016/0022-2836(88)90144-1
- Jamme F, Bourquin D, Tawil G, Vikso-Nielsen A, Buleon A, Refregiers M. 3D imaging of enzymes working in situ. Anal Chem. 2014:86(11):5265-5270. https://doi.org/10.1021/ac403699h
- Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol. 1997:267(3):727–748. https://doi.org/10.1006/jmbi.1996.0897
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021:596(7873):583–589. https://doi.org/10.1038/s41586-021-03819-2
- Larson ME, Falconer DJ, Myers AM, Barb AW. Direct characterization of the maize starch synthase IIa product shows maltodextrin elongation occurs at the non-reducing end. J Biol Chem. 2016:291(48): 24951–24960. https://doi.org/10.1074/jbc.M116.754705
- Liu C, Pfister B, Osman R, Ritter M, Heutinck A, Sharma M, Eicke S, Fischer-Stettler M, Seung D, Bompard C, et al. LIKE EARLY STARVATION 1 and EARLY STARVATION 1 promote and stabilize amylopectin phase transition in starch biosynthesis. Sci Adv. 2023:9(21):eadg7448. https://doi.org/10.1126/sciadv.adg7448
- Malinova I, Mahto H, Brandt F, Al-Rawi S, Qasim H, Brust H, Hejazi M, Fettke J. EARLY STARVATION1 specifically affects the phosphorylation action of starch-related dikinases. Plant J. 2018:95(1):126–137. https://doi.org/10.1111/tpj.13937
- Micsonai A, Wien F, Kernya L, Lee YH, Goto Y, Refregiers M, Kardos J. Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. Proc Natl Acad Sci U S A. 2015:**112**(24): E3095–E3103. https://doi.org/10.1073/pnas.1500851112

- Miles AJ, Janes RW, Brown A, Clarke DT, Sutherland JC, Tao Y, Wallace BA, Hoffmann SV. Light flux density threshold at which protein denaturation is induced by synchrotron radiation circular dichroism beamlines. J Synchrotron Radiat. 2008:15(4):420-422. https://doi.org/10.1107/S0909049508009606
- Miles AJ, Wallace BA. CDtoolx, a downloadable software package for processing and analyses of circular dichroism spectroscopic data. Protein Sci. 2018:27(9):1717–1722. https://doi.org/10.1002/pro.3474
- Miles AJ, Wien F, Wallace BA. Redetermination of the extinction coefficient of camphor-10-sulfonic acid, a calibration standard for circular dichroism spectroscopy. Anal Biochem. 2004:**335**(2):338–339. https://doi.org/10.1016/j.ab.2004.08.035
- Myers AM, Morell MK, James MG, Ball SG. Recent progress toward understanding biosynthesis of the amylopectin crystal. Plant Physiol. 2000:122(4):989–997. https://doi.org/10.1104/pp.122.4.989
- Pérez S, Bertoft E. The molecular structures of starch components and their contribution to the architecture of starch granules: a comprehensive review. Starch Stärke. 2010:62(8):389–420. https://doi.org/10. 1002/star.201000013
- Petoukhov MV, Svergun DI. Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys J. 2005:89(2):1237-1250. https://doi.org/10.1529/biophysi.105.064154
- Pfister B, Lu KJ, Eicke S, Feil R, Lunn JE, Streb S, Zeeman SC. Genetic evidence that chain length and branch point distributions are linked determinants of starch granule formation in Arabidopsis. Plant Physiol. 2014:165(4):1457–1474. https://doi.org/10.1104/pp.114.241455
- Pfister B, Zeeman SC. Formation of starch in plant cells. Cell Mol Life Sci. 2016:73(14):2781–2807. https://doi.org/10.1007/s00018-016-2250-x
- Refregiers M, Wien F, Ta HP, Premvardhan L, Bac S, Jamme F, Rouam V, Lagarde B, Polack F, Giorgetta JL, et al. DISCO synchrotron-radiation circular-dichroism endstation at SOLEIL. J Synchrotron Radiat. 2012:19(5):831–835. https://doi.org/10.1107/ S0909049512030002
- Sawada T, Nakamura Y, Ohdan T, Saitoh A, Francisco PB, Jr., Suzuki E, Fujita N, Shimonaga T, Fujiwara S, Tsuzuki M, et al. Diversity of reaction characteristics of glucan branching enzymes and the fine structure of alpha-glucan from various sources. Arch Biochem Biophys. 2014:562:9–21. https://doi.org/10.1016/j.abb.2014.07.032
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al.

Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012:**9**(7):676–682. https://doi.org/10.1038/nmeth. 2019

- Singh A, Compart J, Al-Rawi SA, Mahto H, Ahmad AM, Fettke J. LIKE EARLY STARVATION 1 alters the glucan structures at the starch granule surface and thereby influences the action of both starchsynthesizing and starch-degrading enzymes. Plant J. 2022:111(3): 819–835. https://doi.org/10.1111/tpj.15855
- Streb S, Delatte T, Umhang M, Eicke S, Schorderet M, Reinhardt D, Zeeman SC. Starch granule biosynthesis in Arabidopsis is abolished by removal of all debranching enzymes but restored by the subsequent removal of an endoamylase. Plant Cell. 2008:20(12): 3448–3466. https://doi.org/10.1105/tpc.108.063487
- Tawil G, Jamme F, Refregiers M, Vikso-Nielsen A, Colonna P, Buleon
 A. In situ tracking of enzymatic breakdown of starch granules by synchrotron UV fluorescence microscopy. Anal Chem. 2011:83(3): 989–993. https://doi.org/10.1021/ac1027512
- Tsai CY. The function of the waxy locus in starch synthesis in maize endosperm. Biochem Genet. 1974:11(2):83–96. https://doi.org/10. 1007/BF00485766
- Waight TA, Kato LK, Donald AM, Gidley MJ, Clarke CJ, Riekel C. Side-chain liquid-crystalline model for starch. Starch Stärke. 2000:52(12): 450–460. https://doi.org/10.1002/1521-379X(200012)52:123.0.CO;2-5
- Wattebled F, Dong Y, Dumez S, Delvalle D, Planchot V, Berbezy P, Vyas D, Colonna P, Chatterjee M, Ball S, et al. Mutants of Arabidopsis lacking a chloroplastic isoamylase accumulate phytoglycogen and an abnormal form of amylopectin. Plant Physiol. 2005:138(1):184–195. https://doi.org/10.1104/pp. 105.059295
- Xie Y, Barb AW, Hennen-Bierwagen TA, Myers AM. Direct determination of the site of addition of glucosyl units to maltooligosaccharide acceptors catalyzed by maize starch synthase I. Front Plant Sci. 2018:9:1252. https://doi.org/10.3389/fpls.2018.01252
- Zeeman SC, Umemoto T, Lue WL, Au-Yeung P, Martin C, Smith AM, Chen J. A mutant of Arabidopsis lacking a chloroplastic isoamylase accumulates both starch and phytoglycogen. Plant Cell. 1998:10(10):1699–1712. https://doi.org/10.1105/tpc.10.10.1699
- Ziegler GR, Creek JA, Runt J. Spherulitic crystallization in starch as a model for starch granule initiation. Biomacromolecules. 2005:6(3): 1547–1554. https://doi.org/10.1021/bm049214p

Séquence peptidique des constructions tronquées PII1 et SS4

PII1-H2H3

LPYSNLGVLESDLEAALVALLKREEDLHDAERKLLSDKNKLRAKEELEKREKTISEASLKHESLQEELKRANVELASQARE IEELKHKLRERDEERAALQSSLTLKEEELEMRQEIANRSKEVSMAISEFESKSQLLSKANEVVKRQEGEIYALQRALEEK EEELEISKATKKLEQEKLRETEANLKKQTEEWLIAQDEVNKLKEETVKRLGEANETMEDFMKVKKLLTDVRFELISSRE ALVFSREQMEEKELLLEKQLEELEEQRKSVLSYMQSLRDAHTEVESERVKLRVVEAKNFALEREISVQKELLEDLREELQ KEKPLLELAMHDISVIQDELYKKANAFQVSQNLLQEKESSLVEAKLEIQHLKSEQASLELLLQEKDEELAEARNKLGEVN QEVTELKALMISREDQLMEATEMLKEKDVHLHRIEGELGSSKLKVTEAEMVVERIAELTNR

PII1∆Nter

SANVLFDKLFARTHRLERQTNQHSVYPDDDDLPYSNLGVLESDLEAALVALLKREEDLHDAERKLLSDKNKLNRAKEE LEKREKTISEASLKHESLQEELKRANVELASQAREIEELKHKLRERDEERAALQSSLTLKEEELEKMRQEIANRSKEVSM AISEFESKSQLLSKANEVVKRQEGEIYALQRALEEKEEELEISKATKKLEQEKLRETEANLKKQTEEWLIAQDEVNKLKEE TVKRLGEANETMEDFMKVKKLLTDVRFELISSREALVFSREQMEEKELLLEKQLEELEEQRKSVLSYMQSLRDAHTEV ESERVKLRVVEAKNFALEREISVQKELLEDLREELQKEKPLLELAMHDISVIQDELYKKANAFQVSQNLLQEKESSLVEA KLEIQHLKSEQASLELLLQEKDEELAEARNKLGEVNQEVTELKALMISREDQLMEATEMLKEKDVHLHRIEGELGSSKL KVTEAEMVVERIAELTNRLLMSTTNGQNQNAMRINNEISIDSMQQPLEKPHDDYGMENKRLVMELSFTRENLRM KEMEVLAVQRALTFKDEEINVVMGRLEAKEQELKKLKEETINDSEDLKVLYALAQERVGEKTMGDLAIEMLQLEAAN LEVEAATSALQKLAKMSTELLTQADMSIEADTTHTVMPERGYSEGSNECLGEVKTEVVRLWSLTEKLLENAGIVAGTS TCMEGVIL

PII1-H1H2H3

SANVLFDKLFARTHRLERQTNQHSVYPDDDDLPYSNLGVLESDLEAALVALLKREEDLHDAERKLLSDKNKLNRAKEE LEKREKTISEASLKHESLQEELKRANVELASQAREIEELKHKLRERDEERAALQSSLTLKEEELEKMRQEIANRSKEVSM AISEFESKSQLLSKANEVVKRQEGEIYALQRALEEKEEELEISKATKKLEQEKLRETEANLKKQTEEWLIAQDEVNKLKEE TVKRLGEANETFMKVKKLLTDVRFELISSREALVFSREQMEEKELLLEKQLEELEEQRKSVLSYMQSLRDAHTEVESER VKLRVVEAKNFALEREISVQKELLEDLREELQKEKPLLELAMHDISVIQDELYKKANAFQVSQNLLQEKESSLVEAKLEI QHLKSEQASLELLLQEKDEELAEARNKLGEVNQEVTELKALMISREDQLMEATEMLKEKDVHLHRIEGELGSSKLKVT EAEMVVERIAELTNR

PII1∆H1

LPYSNLGVLESDLEAALVALLKREEDLHDAERKLLSDKNKLNRAKEELEKREKTISEASLKHESLQEELKRANVELASQA REIEELKHKLRERDEERAALQSSLTLKEEELEKMRQEIANRSKEVSMAISEFESKSQLLSKANEVVKRQEGEIYALQRAL EEKEEELEISKATKKLEQEKLRETEANLKKQTEEWLIAQDEVNKLKEETVKRLGEANETMEDFMKVKKLLTDVRFELISS REALVFSREQMEEKELLLEKQLEELEEQRKSVLSYMQSLRDAHTEVESERVKLRVVEAKNFALEREISVQKELLEDLREE LQKEKPLLELAMHDISVIQDELYKKANAFQVSQNLLQEKESSLVEAKLEIQHLKSEQASLELLLQEKDEELAEARNKLGE VNQEVTELKALMISREDQLMEATEMLKEKDVHLHRIEGELGSSKLKVTEAEMVVERIAELTNRLLMSTTNGQNQNA MRINNEISIDSMQQPLEKPHDDYGMENKRLVMELSFTRENLRMKEMEVLAVQRALTFKDEEINVVMGRLEAKEQE LKKLKEETINDSEDLKVLYALAQERVGEKTMGDLAIEMLQLEAANLEVEAATSALQKLAKMSTELLTQADMSIEADTT HTVMPERGYSEGSNECLGEVKTEVVRLWSLTEKLLENAGIVAGTSTCMEGVIL

SS4∆349

ECTDLWAKVETLQLLLDRATKQAEQAVIVLQQNQDLRNKVDKIEESLKEANVYKESSEKIQQYNELMQHKVTLLEERL EKSDAEIFSYVQLYQESIKEFQETLESLKEESKKKSRDEPVDDMPWDYWSRLLLTVDGWLLEKKIASNDADLLRDMV WKKDRRIHDTYIDVKDKNERDAISAFLKLVSSPTSSGLYVVHIAAEMAPVAKVGGLGDVVAGLGKALQRKGHLVEIILP KYDCMQYDRVRDLRALDTVVESYFDGKLYKNKIWIGTVEGLPVHFIEPQHPSKFFWRGQFYGEQDDFRRFSYFSRA ALELLLQSGKKPDIIHCHDWQTAFVAPLYWDLYAPKGLDSARICFTCHNFEYQGTASASELGSCGLDVNQLNRPDRM QDHSSGDRVNPVKGAIIFSNIVTTVSPTYAQEVRTAEGGKGLHSTLNFHSKKFIGILNGIDTDSWNPATDPFLKAQFN AKDLQGKEENKHALRKQLGLSSAESRRPLVGCITRLVPQKGVHLIRHAIYRTLELGGQFVLLGSSPVPHIQREFEGIEQ QFKSHDHVRLLLKYDEALSHTIYAASDLFIIPSIFEPCGLTQMIAMRYGSIPIARKTGGLNDSVFDIDDDTIPTQFQNGFT FQTADEQGFNYALERAFNHYKKDEEKWMRLVEKVMSIDF SWGSSATQYEELYTRSVSRARAVPNRT

SS4∆403

ESSEKIQQYNELMQHKVTLLEERLEKSDAEIFSYVQLYQESIKEFQETLESLKEESKKKSRDEPVDDMPWDYWSRLLLT VDGWLLEKKIASNDADLLRDMVWKKDRRIHDTYIDVKDKNERDAISAFLKLVSSPTSSGLYVVHIAAEMAPVAKVGG LGDVVAGLGKALQRKGHLVEIILPKYDCMQYDRVRDLRALDTVVESYFDGKLYKNKIWIGTVEGLPVHFIEPQHPSKF FWRGQFYGEQDDFRRFSYFSRAALELLLQSGKKPDIIHCHDWQTAFVAPLYWDLYAPKGLDSARICFTCHNFEYQGT ASASELGSCGLDVNQLNRPDRMQDHSSGDRVNPVKGAIIFSNIVTTVSPTYAQEVRTAEGGKGLHSTLNFHSKKFIGI LNGIDTDSWNPATDPFLKAQFNAKDLQGKEENKHALRKQLGLSSAESRRPLVGCITRLVPQKGVHLIRHAIYRTLELG GQFVLLGSSPVPHIQREFEGIEQQFKSHDHVRLLLKYDEALSHTIYAASDLFIIPSIFEPCGLTQMIAMRYGSIPIARKTG GLNDSVFDIDDDTIPTQFQNGFTFQTADEQGFNYALERAFNHYKKDEEKWMRLVEKVMSIDFSWGSSATQYEELYT RSVSRARAVPNRT

SS4∆466

PVDDMPWDYWSRLLLTVDGWLLEKKIASNDADLLRDMVWKKDRRIHDTYIDVKDKNERDAISAFLKLVSSPTSSGL YVVHIAAEMAPVAKVGGLGDVVAGLGKALQRKGHLVEIILPKYDCMQYDRVRDLRALDTVVESYFDGKLYKNKIWIG TVEGLPVHFIEPQHPSKFFWRGQFYGEQDDFRRFSYFSRAALELLLQSGKKPDIIHCHDWQTAFVAPLYWDLYAPKGL DSARICFTCHNFEYQGTASASELGSCGLDVNQLNRPDRMQDHSSGDRVNPVKGAIIFSNIVTTVSPTYAQEVRTAEG GKGLHSTLNFHSKKFIGILNGIDTDSWNPATDPFLKAQFNAKDLQGKEENKHALRKQLGLSSAESRRPLVGCITRLVP QKGVHLIRHAIYRTLELGGQFVLLGSSPVPHIQREFEGIEQQFKSHDHVRLLLKYDEALSHTIYAASDLFIIPSIFEPCGLT QMIAMRYGSIPIARKTGGLNDSVFDIDDDTIPTQFQNGFTFQTADEQGFNYALERAFNHYKKDEEKWMRLVEKVM SIDFSWGSSATQYEELYTRSVSRARAVPNRT

<i>lc1</i>	Query_7681755		
		_	· · · ·
1c1	Query_7681755	1	MGFS.QAIRLNLASFSSPSPCDYCLTRV
lcl	Query_7681764	1	MGFA.AAFRPNLPAASPLRFRQLSTL
lcl	Query_7681774	1	MAFS.AASRSNLPTSSSQSYGKLCSL
lcl	Query_7681775	1	MGFS.VSVRSSLPLYKLFF.
lcl	Query_7681759	1	MAFS.ARFSSNHLH
lcl	Query_7681766	1	MVFS.AALRPILPATSSHLCSKLCY
lcl	Query_7681770	1	MSFS.AAFQAN
lcl	Query_7681768	1	MGFS.SAILSNLPADT
lcl	Query_7681762	1	MEFAAAAIRFSLPSTSSPLHCDCQCYA
lcl	Query_7681773	1	MGFS.VALRCYLPVESSLHSSKFF
lcl	Query_7681780	1	MGFS.AALRPNLSVTSSLHCSKLGVLDPAISIFAKLNOCFFGLDYTGKISKLTKEMMAKI
lcl	Query_7681782	1	MAFA.AVFHVPPTSSHHYSQLCSL
lcl	Query_7681771	1	MALRASLSFSSIHQOTEFCFA
lcl	Query_7681760	1	MASP.AATHLNFSTSSSISOSORS
lcl	Ouerv 7681761	1	MSMALSASFSLASSHOL
lcl	Query_7681777	1	MALP.ALPRATLSFSSLCOPTEFCFM
lcl	Ouerv 7681763	1	MHHS.SSTHPWLSPLYFALPLLL.
lcl	Ouerv 7681778	1	MALR.ALPRATLSFSFLWOPKECCFM
lcl	Ouery 7681769	1	MTFSASSI, RPTSSPSYSO
lcl	Ouerv 7681758	1	M
lcl	Ouerv 7681765	1	MAVLSPSLOGSFSA
lcl	Ouerv 7681756	1	PVERHOLLKOEMASTPLYGKSFSSSLLRRSHVSEHCAVRL
lcl	Ouerv 7681767	1	MNVDISMRFOESIVAAVAVVIWFTVPEKHITKCKL
lcl	Ouerv 7681757	1	MFRGHAPLHRLPSPPPPPAAAAGALPSASPSCRTSTHVP
lcl	Ouery 7681779	1	M
lcl	Ouerv 7681783	1	MPLSSTTSPSAGAAAAA.AVRTASPPRRIATHVL
lcl	Ouery 7681772	1	MPPLSPSSSP
lcl	Ouery 7681781	1	MRLSIGSPSPSPPPAVAA.ALRSTPPSRTASHVM
lcl	Query_7681776		

1c1	Query_7681755		
lcl	Query_7681755	28	VNHKQKSLVAFPSIT.RRKRHLLLSVOSVL.HNTRPN.IND
lcl	Query_7681764	26	RHSWKQKTLP.FVALT.RGKGHSLLIVESVL.NNSKSS.IND
lcl	Query_7681774	26	RFSRKQNKVA.FLTTT.KRKGSSLRIIR <mark>SV</mark> L.NNRK <mark>S</mark> S.ISG
lcl	Query_7681775	18	SPRLNSRQNRVDCITTISKRRKSSPLQIVKSVL.NSSNSS.IDD
lcl	Query_7681759	14	.LNPNPKVHWKHKLPGRYVTSG.KRRVRSLGLVRAVLPDGKKSS.VNG
lcl	Query_7681766	25	SRLNKRQDKLAFPMTKRRGHCLKVVRAVL.NKRKIS.IND
lcl	Query_7681770	12	STSLPHRHSKLSFLRLVRKQKKSAIIGTRNGKKCSIQMTRSIL.GNRTSC.IDN
lcl	Query_7681768	21	RLRFNVKQNRVNFITTGKIIRPSKWIVK <mark>SV</mark> L.NCSK <mark>S</mark> S. <mark>I</mark> ND
lcl	Query_7681762	28	RLRRKQQKLPFLSTTL.TWKNHSLYCVK <mark>SV</mark> L.NNRN <mark>S</mark> R. <mark>I</mark> SD
lcl	Query_7681773	23	SPRFNLKQNRLAFITTSKIESPSLQIVK <mark>S</mark> IS.NNMN <mark>S</mark> S. <mark>I</mark> NE
lcl	Query_7681780	60	VLSSRPTRDWKQKRLP.LLAVT.KRRGYSLFIVK <mark>S</mark> II.NSSK <mark>S</mark> S.VND
lcl	Query_7681782	24	GLNRKQKRLA.VMTTS.KRKGHSRRIVK <mark>SV</mark> L.NNRK <mark>S</mark> S. <mark>I</mark> ND
lcl	Query_7681771	22	RPKWKKKRLVLMTAHRGPPSSRIVR <mark>SV</mark> L.DNRK <mark>S</mark> N. <mark>I</mark> KG
lcl	Query_7681760	24	SLRFGRNRTNFFYSTNQKRRSHSLKVVQ <mark>SV</mark> L.NNCK <mark>S</mark> N.LN.D
lcl	Query_7681761	25	LKWKQTRLEFLKPHRRNSQYLWIVK <mark>SV</mark> L.NRR <mark>S</mark> SSVSE
lcl	Query_7681777	26	RLEWKK.RLVLMTAHHGRGPSSRIVR <mark>SV</mark> L.DNRK <mark>S</mark> N. <mark>I</mark> TG
lcl	Query_7681763	26	
lcl	Query_7681778	26	RLEWKK.RLMLMTAYHGRGPSSRIVR <mark>SV</mark> L.DNRK <mark>SN.I</mark> TG
lcl	Query_7681769	47	VKSVL.NDNRP.SF.N
lcl	Query_7681758	2	PSSRILRSVL.DNRKSN.INR
lcl	Query_7681765	20	PVVRSSLFKQRSFRCQKRGHS.AWIIRAVMNKESS.IND
lcl	Query_7681756	41	
TCT	Query_7681767	36	WIIKSVINNKSING
TCT	Query_7681757	40	FRPKLSFMVAFQAQHVKYAPNLIKSVV.KSLRSN.IIDGD
TCT	Query_7681779	34	FRQKLGFLAAFQAQRVKCSPHLIRSIV.KSSRLD.INDGD
TCT	Query_/681/83	34	FRQKLGIPAGFQAQHVKCLPHLIKSIV.RGARSD.IIDGD
TCT	Query_/6817/2	33	FRQKLSFMVSFQAQHMKCAPHLIKSVV.KGIRAN.ITDGE
lcl	Query_7681781 Query_7681776	35	

						α1							
1c1	Query_7681755			. ووور	عففة	يعععه	ععع	ll			leele	بععععه	معععععع
lcl	Query_7681755	66	NGSA	SANV	FDKI	FARTH	RLE	RQTN	QHSVY	PDDDLP	YSNLGV	LESDLE	AALVALLK
lcl	Query_7681764	64	NGAA	SAKV	LERI	FAQTQ	KLE	QGIS	RDGEP	LQDFHLA	L.DLQT	LESDLLJ	AALTALKQ
lcl	Query_7681774	64	NGAS	PARI	LERI	FAQTQ	KLE	EQMN	RNSHH	PQDIQLG	F.NLEI	LESDLHI	AALAALKK
lcl	Query_7681775	60	NGAT	PARI	LERI	FAQTQ	ΚLΕ	КQМG	PHSHL	PADVYPG	F.NLEI	LESDLLI	AVLEALRK
lcl	Query_7681759	59	YGLG	PARI	LERI	FAQTQ	KLE	ERMS	RDSGV	GKDVQFG	L.NLEI	LESDLQI	AVLAALKK
lcl	Query_7681766	63	NGPT	SAKI	LERI	FEQTQ	ΚLΕ	EQMI	QDSYP	PEDVQPL	L.NLEV	LESDLQI	AVLLALKQ
lcl	Query_7681770	64	NGNGAT	PARV	LERI	FAQTQ	KLE	EQIS	RDSSF	SKDIQLG	L.NLET	LESDLQI	AVLTALRK
lcl	Query_7681768	61	NGAT	PARI	LERI	FAQTK	ΚLΕ	ЕЕМІ	SDSQL	PRDVQPG	F.NLEI	LESDLLI	AVLDALRK
lcl	Query_7681762	67	NGAA	PARV	LERI	FARTH	ΚLΕ	E.MN	KASSV	PQDVQLD	L.NFEV	LESDLH?	AALASLKK
lcl	Query_7681773	63	NGAT	PARV	LERI	FAQTL	ΚLΕ	EQMS	RSSRL	PEDVQPV	V.NLEI	LESDLL	ALLKALKK
lcl	Query_7681780	104	NGAT	PARI	LERI	FAQSQ	ΚLΕ	QGMS	RDAQP	PKDFHLF	L.NLET	LESDLQ/	AALTALKQ
lcl	Query_7681782	62	NGST	PARV	LERI	FAQTQ	ΚLΕ	EHMS	RDPGL	PLDIQLG	L.NLET	LESDLQI	AALVALKK
lcl	Query_7681771	59	NGAT	PARV	LGR	FAQTQ	ΚLΕ	EQIG	RNPYL	PQVAELG	L.DLGE	LESDLLI	DALAALKR
lcl	Query_7681760	64	NGAN	EAKL	LER	YAQTQ	RLE	EHVS	KDPHF	PQDVWLG	L.SLEN	LESDLQ/	AALAVLKK
lcl	Query_7681761	63	DGAT	PARI	LEKI	FDQTQ	ΚLΕ	EQIG	RDPYS	PHVAELG	F.SLGK	LESDLQ	IALAALKN
lcl	Query_7681777	63	DEAT	PARV	LERI	FAQTQ	ΚLΕ	QQIG	RNIYF	PQVAELG	L.NLGK	LESDLQI	DALAALKK
lcl	Query_7681763	34	DY.GAA	SARV	FER	FTQ		NRMI	GD	EP.	DLRI	LESDLK	AALAALKM
lcl	Query_7681778	63	EEET	PARVI	LERI	FAQTQ	ΚLΕ	QQIG	RNIYF	PQVAELG	L.NLGK	LESDLLI	DALAALKK
lcl	Query_7681769	65	NY.GAP	SARL	ER	FEQTQ	ΚLD	NRMI	GE	EP.	DLRD	FESDLL	SALMELKE
lcl	Query_7681758	33	NEAT	PARV	LERI	FAQTQ	ΚLΕ	QQIG	RNPYL	PQVAELG	L.NIGK		DALAALK.
lcl	Query_7681765	57	DEAI	PTRI	LERI	FVQTQ	KLE	ΕΚΙG	KDLNL	PQDIQLE	L.NLGT	LESDLQI	EALTVIRK
lcl	Query_7681756	78	NGHEPS	SARV	LERI	FAQTQ	KLE	EQMG	KRRRS	LKDIKLD	I.NLEV	L E S D L Q I	AALAALKK
lcl	Query_7681767	76	DEAI	PARI	LERI	FVQTQ	KLE	ΕΚΙΝ	KNSNP	PQDIELE	H.YLGK	LESDLQ	IALTVLRK
lcl	Query_7681757	78	NGMT	PAREI	LERI	FAKTQ	SL.		.DTSA	SHDSELS	M.SIEV	L K S E F E J	RALSILRK
lcl	Query_7681779	72	NGTT	PARE	LERI	FAKTK	SL.		.DPGA	SQDRELS	M.SIEV	LKTEFE	AALSILRK
lcl	Query_7681783	72	NGTT	PARE	LERI	FAKTK	SL.		.DPSA	SQGRELS	M.SIEV	LKTEFE/	AALSILRK
lcl	Query_7681772	71	NGAT	PAREI	LERI	FAKTQ	RL.		.DTSA	SQDSELS	M.SIDV	LKSEFE/	AALSTLRK
lcl	Query_7681781	73	NGTT	PARE	LERI	FARTQ	SL.		.DTGA	SHDSELS	V.SIEV	LKSEFE(GALSILRN
lcl	Query_7681776	24	N E T	PARE	LERI	FAKTK	SL.		.DPSA	SQDGELS	M.SIEV	LKTEFE	A A 🖸 S I 🖬 R K

1 <i>c</i> 1	Query_7681755			ll
lcl	Query_7681755	124	REEDIHDAERKLLSDKNKLNRAKEEEKREKTISEASLKHESLQEELKRANVELASQA	RE
lcl	Query_7681764	121	KEDDLQDAEKMVVLEQSELSRAKDELEQREKEIAAASSKHEKLEKLTQANLAFASQA	SQ
lcl	Query_7681774	121	KEEDLQDAERTVFFEHCELHRTKEE T EQREKEIAAASCRYEKIGEELKQANLGLASQA	RΗ
lcl	Query_7681775	117	KEEDLQDAERQVLSEHSDLNHAKEMTELRENEIAIAYSKHEKLEGELKLANVYLASQS	RQ
lcl	Query_7681759	116	KEEDLEDAERRVCLEHSELNRAKEELLRREREIDVACSRHEKLEEELGQSNLKLVSQA	RH
lcl	Query_7681766	120	KEEDLQDAESIVLLEHSQLNCAKEELERRDEEIAAARCKHEQLEEEMKKSTRTLASQA	RQ
TCT	Query_7681770	123	KEED LODAERKVLFEHAKLNHTKQEFEKREEAITSALCRQEKLEEELKQANNN LASOA	RQ
TCT	Query_7681768	118	KEED LOHAEKOVLSEHNDLNFAKOETOOREKEIAVAHSKHEKLEGELREAN LNLASOA	RQ
101	Query_/681/62	120	KEED LEEAEKKLLLD SELNCAIEEEKKKEEVIAAALSKQELLEWDLKQANLKFASQA	DF
101	0uery 7681780	161	KED LODAERMYLLEOSOLTPAKDETEODENETIAAISSKDEKLEEELKOANLASSOL	CO.
101	0uery 7681782	119	KEEDLODALGMVIMEHTELNRAKEETKRHAEETAVACSKHEKLEETIKOANINILASRA	RO
lcl	Ouerv 7681771	116	KEEDLODTERKVLMEYNEVNHAKLELEOREEEIAAASSROEKLEDELMOANLTLASOA	AĒ
lcl	Ouerv 7681760	121	KEEDLODAERTILLERSOLNNAREKTEKOEEEITYAYRKOOELEDELKEANLNLVSOT	RL
lcl	Query_7681761	120	KEEDLQDAEKKLLLEYNEINLARKDLERREKEIAAANMKQEKLEGELRLANLDLASQA	ΑE
lcl	Query_7681777	120	KEEDIQDTERKVLMEYNELNRAKIELEQRVEEMAAANSRQEKLENELRQANLILVSQA	ΑE
lcl	Query_7681763	78	KEDHLMEAERMVLLENSKLKLTKEE T ERQESEIEAARVRYEKLEEEMKETTVKLVSQA	GQ
lcl	Query_7681778	120	KEDDIQDTERKVLMEYNELNRAKIELEQRVEEMEAANSRQEKLENELRQANLVLVSQA	ΑE
lcl	Query_7681769	113	KEDHLQEVERTVLLENGKLKDAKEE T ERQEGEIKAAREKYERLEDEMKEAMASLVSQA	GQ
lcl	Query_7681758	89	KEEDIQDTERKVSMEYNELNHAKLEEEQRAEEIASASSRQEKLENENLTLASQA	ΕE
lcl	Query_7681765	114	KVEN LEDAENKLLSEYRELREAKEDIRKREAVIESTLLRQEMLENELKQANLDLAFQA	TE
TCT	Query_7681756	137	KEEDLUYAEKMULMDRARLTDVKQDFDHREEEIIAAHAKQAKLEEDLKKAQEDFASQE	κQ
101	Query_/681/6/	120	KEED LEAAENKIS LEIKULNNAKNEINKKEENIS DAFLKQELLENEIN LANDULAS KA	ILL
101	Query_7681737	123	KEKNIDAEKKVSUDDSELNOKODDODEEDIIKAVSODHEKAIMKASDDIIIO	L D L D
	Query 7681783	123	KEKD LRD AEKKVS VD R SRLNO TKOD DO REED LIKAY SROHEMEKALMKASRDLTLOV	RO
lcl	Ouerv 7681772	122	KE B D L B D A E N B VS V D O V B L N B A K K D D O B E B G I N B A Y A B O O E M E B S L G K A S B D L V L O V	RÔ
lcl	Query_7681781	124	KERDLRSAEKRVSDDRIRLSKTKODDOREEAIRKAYVROOGIEKALKKASRDLALRV	κõ
lcl	Query_7681776	74	KERDLWDAEKKVSIDRSRLNQTKQDDDQREEDIIKAYSRQHEMEKALMKASRDLSLQV	Ϋ́RQ
lcl	Query_7681755		α3	20
lcl	Query_7681755	194		ee KA
<i>lc1</i> lc1 lc1	Query_7681755 Query_7681755 Ouery_7681764	184 181	α3 000000000000000000000000000000000000	L KA
<i>lcl</i> lcl lcl lcl	<i>Query_7681755</i> <i>Query_7681755</i> <i>Query_7681764</i> <i>Ouery_7681774</i>	184 181 181	α3 20202020202020202020202020202020202020	C C C C C C C C C C C C C C C C C C C
<i>lcl</i> lcl lcl lcl lcl	Query_7681755 Query_7681755 Query_76817764 Query_7681774 Query_7681774	184 181 181 177	α3 ΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩ	KA EA EA
<i>lcl</i> lcl lcl lcl lcl lcl	Query_7681755 Query_7681755 Query_7681774 Query_7681775 Query_7681775	184 181 181 177 176	α3 ΔΟΔΟΔΟΔΟΔΟΔΟΔΟΔΟΔΟΔΟΔΟΔΟΔΟΔΟΔΟΔΟΔΟΔΟΔΟ	KA EA EA EA
1c1 1c1 1c1 1c1 1c1 1c1 1c1	Query_7681755 Query_7681755 Query_7681754 Query_7681774 Query_7681775 Query_7681756 Query_7681766	184 181 181 177 176 180	α3 000000000000000000000000000000000000	KA EA EA EA
1c1 1c1 1c1 1c1 1c1 1c1 1c1	Query_7681755 Query_7681755 Query_7681764 Query_7681774 Query_7681775 Query_7681759 Query_7681776 Query_7681770	184 181 177 176 180 183	α3 OCOOCOOCOOCOOCOOCOOCOOCOOCOOCOOCOOCOOCO	KA EA EA EA EA
1c1 1c1 1c1 1c1 1c1 1c1 1c1	Query_7681755 Query_7681755 Query_7681764 Query_7681775 Query_7681775 Query_7681776 Query_7681776 Query_7681776 Query_7681776	184 181 177 176 180 183 178	α3 20202020202020202020202020202020202020	KA EA EA EA EA
1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1	Query_7681755 Query_7681755 Query_7681754 Query_7681774 Query_7681775 Query_7681775 Query_7681766 Query_7681770 Query_7681762 Query_7681762	184 181 177 176 180 183 178 183	C3 OCOOCOOPOOLOOCOOPOOLOOCOOPOOLOOCOOPOOLOOCOOPOOLOOCOO TEELKHKLRERDEERAALOSSITIKEEELEKMRQETANRSKEVSMAISEFESKSQLT TEDLKLQLKEQDQKVAAAQSAISAREDEMDKMRHELVKKTEEAEKIRSELTSKSQLT TEDLKLQVKEREDVIFAAKSAISLKEEELDKMRNELLLKSEEAAKTESELKSKSHLT TEDLKLQVKEREDVIFAAKSAISLKEEELDKMRNELLKKSEEAAKTESELKSKSHLT TEDLKLQVKEREDVIFAAKSAISLKEEELEKMRSKLTKKSEEAAKTESELKSKSAQMIN TEDLKLQVKEREDVIFAAKSAISLKEEELEKMRSELLKKSEEAAKTESELKSKJON TEDLKLQVKEREDVIFAAKSAISLKEEEMEKLRVELVKKIEEAATISELKSKJON TEDLKLQVKEREDVATOSISIKEEEMEKLRVELVKKIEEAATIACDLQSRDQLIS TEDLKLQVKEREEEVAATQSAISLKEEEMEKLRVELVKKIEEAATIACDLQSRDQLIS TEDLKLQVKEREEEVAAGQSAILIKOHETEKKISELTKKSEEVAKMDSELQYKAQUTI TEDLKLQLKEREEEVAAGQSAISLKEEETEKKISELTKKSEEVAKMDSELQYKAQUTI	KA EA EA EA EA EA
1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1	Query_7681755 Query_7681755 Query_7681754 Query_7681774 Query_7681775 Query_7681756 Query_7681766 Query_7681762 Query_7681762 Query_7681773 Ouery_7681773	184 181 177 176 180 183 178 183 183 221	α3 20202020202020202020202020202020202020	KA EA EA EA EA EA EA EA
1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1	Query_7681755 Query_7681755 Query_7681754 Query_7681774 Query_7681775 Query_7681776 Query_7681768 Query_7681768 Query_7681773 Query_7681780 Ouery_7681780	184 181 177 176 180 183 178 183 183 180 221 179	α3 COCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCO	KA EA EA EA EA EA EA EA EA EA
1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1	Query_7681755 Query_7681755 Query_7681764 Query_7681774 Query_7681775 Query_7681776 Query_7681768 Query_7681770 Query_7681772 Query_7681773 Query_7681780 Query_7681782 Query_7681782	184 181 177 176 183 178 183 180 221 179 176	C3 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CC KA EA EA EA EA EA EA EA EA EA EA
1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1	Query_7681755 Query_7681755 Query_7681774 Query_7681775 Query_7681775 Query_7681776 Query_7681762 Query_7681762 Query_7681773 Query_7681782 Query_7681771 Query_7681771	184 181 177 176 180 183 183 180 221 179 176 181	C3 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	C KA E E E A E E A E A A A A A A A
1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1	Query_7681755 Query_7681755 Query_7681754 Query_7681774 Query_7681775 Query_7681776 Query_7681766 Query_7681768 Query_7681762 Query_7681782 Query_7681782 Query_7681771 Query_7681761	184 181 177 176 180 183 178 183 221 179 176 181 180	C3 COCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCO	KA EA EA EA EA EA EA EA EA EA EA EA EA
1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1	Query_7681755 Query_7681755 Query_7681754 Query_7681774 Query_7681775 Query_7681776 Query_7681766 Query_7681762 Query_7681773 Query_7681773 Query_7681771 Query_7681771 Query_7681771 Query_7681771 Query_7681771	184 181 177 176 183 178 183 183 183 183 221 179 176 181 180 180	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	KA EA EA EA EA EA EA EA EA EA EA EA EA EA
1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1 1c1	Query_7681755 Query_7681755 Query_7681754 Query_7681774 Query_7681775 Query_7681775 Query_7681762 Query_7681762 Query_7681762 Query_7681773 Query_7681780 Query_7681771 Query_7681761 Query_7681771 Query_7681761 Query_76817771	184 181 177 176 180 183 183 180 2219 176 181 180 1380 1380	C3 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	KA EEA EEA EEA EEA EEA EEA EEA EEA EEA E
lcl	Query_7681755 Query_7681755 Query_7681774 Query_7681775 Query_7681775 Query_7681770 Query_7681762 Query_7681762 Query_7681782 Query_7681782 Query_7681771 Query_76817761 Query_76817761 Query_7681777 Query_7681777	184 181 177 176 180 183 178 180 221 176 181 180 138 180	C3 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	KA EEA EEA EEA EEA EEA EEA EEA EEA EEA E
lcl	Query_7681755 Query_7681755 Query_7681774 Query_7681774 Query_7681775 Query_7681776 Query_7681766 Query_7681762 Query_7681762 Query_7681780 Query_7681782 Query_7681781 Query_7681761 Query_7681761 Query_7681776 Query_7681778 Query_7681778 Query_7681776	184 181 177 176 180 183 183 183 183 183 183 183 180 180 180 180 145	C3 COCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCO	KAA EEEAA EEEAA EEAAAAAAAAAAAAAAAAAAAAA
lcl	Query_7681755 Query_7681755 Query_7681754 Query_7681774 Query_7681775 Query_7681776 Query_7681766 Query_7681762 Query_7681773 Query_7681780 Query_7681771 Query_7681771 Query_7681777 Query_7681777 Query_7681775 Query_7681775 Query_7681775 Query_7681758 Query_7681758	184 181 1776 180 183 183 183 183 183 180 176 180 1380 1380 173 145	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	KAA EEAA EEAA EEAAA EEAAAA EEAAAAAA EEAAAAAA
lcl l	Query_7681755 Query_7681754 Query_7681754 Query_7681775 Query_7681775 Query_7681770 Query_7681762 Query_7681762 Query_7681773 Query_7681782 Query_7681781 Query_76817761 Query_76817761 Query_76817773 Query_76817763 Query_7681778 Query_7681758 Query_7681755 Query_7681755	184 181 177 176 180 183 178 180 2219 176 181 180 138 180 138 180 138 180 173 145 174	C3 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	C KA EAA EAA EAA EAA EAA EAA EAA
1c1	Query_7681755 Query_7681755 Query_7681774 Query_7681775 Query_7681775 Query_7681770 Query_7681762 Query_7681762 Query_7681782 Query_7681782 Query_7681771 Query_7681771 Query_7681773 Query_7681773 Query_7681778 Query_7681778 Query_7681775 Query_7681755 Query_7681756	184 181 176 180 183 178 183 183 180 179 176 181 180 180 138 180 173 145 174 193	C3 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	KAAAEEEAA EEEAAEEEAAEEAAEETEETEKETETE
1c1	Query_7681755 Query_7681755 Query_7681754 Query_7681774 Query_7681775 Query_7681776 Query_7681766 Query_7681762 Query_7681762 Query_7681782 Query_7681782 Query_7681771 Query_7681761 Query_7681763 Query_7681763 Query_7681755 Query_7681756 Query_7681767 Query_7681767	184 181 176 180 183 178 183 183 180 173 180 180 180 173 145 174 197 193 189	C3 COCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCO	Q KAA KAA EE EE EA A EE A
lcl	Query_7681755 Query_7681755 Query_7681754 Query_7681775 Query_7681775 Query_7681770 Query_7681760 Query_7681760 Query_7681780 Query_7681771 Query_7681771 Query_7681773 Query_7681773 Query_7681773 Query_7681775 Query_7681758 Query_7681756 Query_7681757 Query_7681757 Query_7681757 Query_7681757	184 181 1776 180 183 183 183 180 176 180 1380 176 180 1380 173 145 197 193 183	C3 20202020202020202020202020202020202020	KEAA EEAA EEAAAAA EEAAAAAA EEAAAAAA EEAAAAAA
1c1	Query_7681755 Query_7681754 Query_7681754 Query_7681775 Query_7681775 Query_7681776 Query_7681762 Query_7681762 Query_7681780 Query_7681780 Query_7681781 Query_7681771 Query_7681761 Query_76817761 Query_76817763 Query_76817765 Query_7681755 Query_7681755 Query_7681757 Query_7681757 Query_7681757 Query_7681757 Query_7681757 Query_7681757	184 181 177 176 180 178 183 178 183 221 179 181 180 138 180 138 180 174 193 189 183 183	C3 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	KAA EEAA EEAAA EEAAAA EEAAAA EEAAAA EEAAAA EEAAAA EEAAAA EEAAAA EEAAAA EEAAAA EEAAAA EEAAAA EEAAAA EEAAAA EEAAAAA EEAAAAA EEAAAAA EEAAAAA
lcl l	Query_7681755 Query_7681755 Query_7681754 Query_7681774 Query_7681775 Query_7681776 Query_7681762 Query_7681762 Query_7681780 Query_7681780 Query_7681781 Query_7681771 Query_7681761 Query_7681773 Query_7681773 Query_7681775 Query_7681755 Query_7681757 Query_7681757 Query_7681757 Query_7681757 Query_7681757	184 181 176 180 183 178 183 183 180 179 1761 180 180 175 180 174 197 189 183 183 183 183	C3 C1000000000000000000000000000000000000	KEEAA EEEAA EEEAAAEEAAAEEAAAEEAAAEE KECEAAAEEAAAE
1c1 1	Query_7681755 Query_7681755 Query_7681754 Query_7681774 Query_7681775 Query_7681776 Query_7681766 Query_7681762 Query_7681762 Query_7681782 Query_7681781 Query_7681781 Query_7681763 Query_7681763 Query_7681763 Query_7681755 Query_7681755 Query_7681755 Query_7681777 Query_7681777 Query_7681773 Query_7681773 Query_7681773 Query_7681773 Query_7681773 Query_7681773	184 181 176 180 183 178 183 183 180 173 145 180 180 173 147 197 193 183 183 183 183 183 184 184	C3 COCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCO	KEAA EEAA EEAA EEAA EEAA EEAA EEAA EEAA

	-			•		•			•			•			•			•
lcl	Query_7681755	244	NEVVKR	QEGEI	ΥA <mark>L</mark>	QRAL	EEKE	Ε <mark>Ε</mark> Ι	ΕIS	KATK	KLE	QEK	LREI	EAN	L K K Q	TEE	ΙLΙΑ	QDEV
lcl	Query_7681764	241	NEVMKK	QEIEL	QE 🗖	REAI	WERE	ΕEL	ΕTS	LTQR	KLE	EEK	L K V A	EAK		TME	LLA	QEEL
lcl	Query_7681774	241	NEVVNR	QAVEV	QG <mark>L</mark>	RKSL	QEKE	ΕEL	EVS	QMQR	KLE	VEK	L K V A	EEK	EKQ	ΤΜΕ	LLA	QEEL
lcl	Query_7681775	237	NEVVKK	QEIEL	QQ <mark>L</mark>	KNAI	RDKQ	ΕKL	EVS	KTLR	KLE	EEK	L K V A	EAN	EKQ	ΤΜΕ	IL I A	QEEL
lcl	Query_7681759	236	NEVVKK	QETEI	QS <mark>L</mark>	RKVI	QEKE	ΕEL	EAS	VALR	KVE	EEK	LKVV	EAN	EKR	TME	LLS	QDAL
lcl	Query_7681766	240	NEVIQK	QEYEL	QG <mark>L</mark>	REAI	REKE	ΕEL	EVS	LTLK	KLE	EEK	L <mark>K</mark> A A	EAN	EKR	ΤΜΕ	LLA	QEEL
lcl	Query_7681770	243	NEIIRK	QKVEI	QE <mark>L</mark>	QKLI	REKV	QΕL	EAS	VTLK	KIE	EEK	LKLA	EAN	KKQ	ΤVΕ	ILSA	QEEL
lcl	Query_7681768	238	NEVVKK	QEIEI	QR <mark>I</mark>	KKGL	QEKE	ΚEL	EVS	VALR	KVE	EEK	LKVV	νQ T N	EKQ	AME	LIA	QEGL
lcl	Query_7681762	243	NEVIKK	QELEL	QEL	RRAV	EEKE	ΕEL	EFS	MTLK	KAE	EDK	V <mark>K</mark> SM	IEAT	EER	TRE	LLA	QEDL
lcl	Query_7681773	240	SEVVKR	QEIEL	QG <mark>L</mark>	QMLI	REKE	ΕEL	EVS	TNLR	KFE	EEK	L K V V	E S N	EDR	TRE	LLI	QEGL
lcl	Query_7681780	281	NEVLKK	QKIEL	QG <mark>L</mark>	KEAI	REKD	ΚQL	ΕTS	MTLR	KLE	EEK	L <mark>K</mark> A A	EAK		ΤΜΕ	LLA	QEEL
lcl	Query_7681782	239	NEVVKK	QEIEL	QEL	QKSI	QEKE	ΕEL	EES	MMLR	KLE	EKK	L K V A	EAN	EKK	TMD	LLA	KEEL
lcl	Query_7681771	236	NEVVQR	QEVEL	QNL	RRKI	QEKE	ΚEL	EDF	LTMQ	KTE	EEK	L K V A	KSN	EKQ	AMD	ILVA	KKEM
lcl	Query_7681760	241	NEVVKR	QEVEL	QΜ <mark>ι</mark>	KKTV	VEKE	ΚEF	ELS	VKLQ	KLE	VER	LEVV	ΈKΝ	EKR	ΤΜΕ	LLA	QEEL
lcl	Query_7681761	240	NKIVEK	QVVEL	QEL	QGAI	QKKD	DEL	EMS	ISVQ	KSD	AEK	L K V A	EAK	EKQ	ΤΤD	ΊLVA	QVEL
lcl	Query_7681777	240	NEVVQR	QEVEL	QNL	QREI	QEKE	ΚEL	QVF	LTMQ	KTE	EEK	LKVS	KSN	EKQ	AMD	IL I A	KQEM
lcl	Query_7681763	198	NEVMKK	QEAEL	ΕE	KRAV	REKE	ΚEΙ	ΕVL	LVER	EVE	REK	LRVA	EAN	EKQ	AMD	MLA	QEEL
lcl	Query_7681778	240	NEVVQR	QEVEL	QNL	RREI	QEKE	ΚEL	QVF	LTMQ	KTE	DEK	LKVS	KSN	EKQ	AMD	LIA	KQEM
lcl	Query_7681769	233	NEVVKK	QEIEL	QE <mark>L</mark>	RSVV	QQRE	ΕEL	RLS	VAAR	DVE	GEK	L K V A	EAS	EKQ	AME	LLT	QEEL
lcl	Query_7681758	205	NELVQR	QEVEL	QNL	RIEI	QEKE	ΚEL	EVF	STMQ	QTD	EEK	LKVS	KSN	AKH	AMD	ILVA	KKEM
lcl	Query_7681765	234	KEIVNK	QTLEI	EEL	KKAV	EEKD	ΕEV	RIA	MALF	ΕIΕ	EEK	L K V V	E G N	EKQ	TLD	LLA	HEEL
lcl	Query_7681756	257	NKVIRK	QE A E V	QDL	KEAI	MSKE	ΕEL	ADT	IQQR	KND	EEK	L K N A	KAN	EQR	AVE	LSS	QEEL
lcl	Query_7681767	253	NEILKK	QEVEI	ΕE	KETI	REKN	ΕEL	EIS	TMLL	ESE	NEK	LKVV	EEN	EKQ	ΤMD	ILVA	NEEM
lcl	Query_7681757	249	NQDVVQ	QEATV	R E 💶	RSET	EKKA	ΙDΙ	AIS	NELR	KAN	EEK	LKIA	EQE	EKQ	NLG	LAA	QQEL
lcl	Query_7681779	243	NQAIAR	QEATV	REL	QSEI	KRKE	ΤDΙ	ERL	NDLT	KAN	EEK	L <mark>K</mark> V A	EQD	EKQ	NSG	IAA	QQEL
lcl	Query_7681783	243	NQAIAR	OEATI	RE	RSET	KRKE	ΤEV	ERL	NELA	KAN	EDK	LKFA	EQE	EKQ	NSG	IAA	QQEL
lcl	Query_7681772	242	NQAAEQ	QEATI	КЕ <mark>L</mark>	RSEI	KRKE	IDF	SRS	NELR	KAN	EQK	LKIA	EQE	ERQ	NMG	LAA	QKEL
lcl	Query_7681781	244	NQAIAQ	O EATV	REL	QSEI	KR <mark>K</mark> T	MDI	ARS	NESR	KTN	EEK	LKVA	EQE	EKQ	SLG	LAA	QQEL
lcl	Query_7681776	194	NOALAR	OEATI	REL	OSEI	KRKE	ΤΕΙ	ASS	NELR	KAN	EEK	LKVV	EKE	ЕКО	NLG	IAA	OOEL
						-		_										

1c1	Query_7681755		عفعفعفع	مععععه	معيعف	لععععه	مقععععمه	ممتحمحمحمحمحم
lcl	Query_7681755	304	NKLKEETVKRLGEANE	TMEDFMK	VKKLLT	DVRFEL	ISSREALVE	SREQMEEKE L LLE KOL
lcl	Query_7681764	301	KKLAEQASRHMGEANE	TFKDFTR	VKQLLS	DVRSEL	VSSQKSLAS	SRQQMEQQEQLLKMQL
lcl	Query_7681774	301	K K L A E E A S R H A G E T N E	TLEDFRR	VKKLLA	DVRSEL	VF <mark>S</mark> QKSLAS	SRQKMEEQEKLLETQW
lcl	Query_7681775	297	K K L A D N A S K Q I V E T K E	TMENFRR	VKKLLI	DVRSEL	VSSQKSLAS	SRKRMEEQEKLLKQQL
lcl	Query_7681759	296	K K L A E E A S R R M E E T N D	TLEDFRR	VKKLLS	DVRSEL	VSSQKSLAS	SRKQMEEQEHLLGKQL
lcl	Query_7681766	300	NKVAEKA <mark>SK</mark> HARETDE	TLEDFRR	VKKLLG	DVRSEL	VS <mark>S</mark> QKSLAS	SRQKMEEQEVQLEKQL
lcl	Query_7681770	303	EKLREEAAKHMRDANE	TLEDFRR	VKKLLV	D V RSE L	MS <mark>S</mark> QKSLAS	SRRKMEEQEHQLERQL
lcl	Query_7681768	298	K R L A N E T S K R M V E T N E	TMKDFKR	VKKLLV	D V RSE L	VSSQKSLAF	SRKRMEEQDLLLKKQL
lcl	Query_7681762	303	KKLSDEASKHAAESNE	TLEDFRR	VKRLLG	DVRAEL	ISSQQSLVS	SRQRMEEKEKLLEKQI
lcl	Query_7681773	300	NKLAKEASKQVRDTNE	ALEDFGR	VYKLLE	DVRSEL	ISSQKSLAF	SRKQMEEQEQLLKTQL
lcl	Query_7681780	341	KKLAEEASRHTGEANE	TFEDFRR	VKQLLS	DVRSQL	VSSQKSLAS	SRQQMAQQEQLLEKQL
lcl	Query_7681782	299	KKLAEDAAKHMGESNK	TMKEFRR	AKRLLH	DVRSEL	VSSQKSLAS	RQKMQEQEKLLEKQL
Tet	Query_7681771	296	KKLEEETSKYGGEANE	TLEDFRR	VKKLLA	DVRSEL	TSSQRALTS	SREKMEEQENLLEERL
TCT	Query_/681/60	301	KKIKKEASKKIVEMNK	IVNDFNR	VKKLLA		VSSQKSLVS	SRKKIEEQEDILEROM
101	Query_/681/61	300	KKLAGEISKHVGDANE	A. EFGI	VERLIS	DVRSDL	ISSQKALAS VCCODALTC	SRUKIESUDULLEMUL
101	Query_/681///	300	KRLEVEISNIGGEANR	SLEDF KK	VKKLLA	DVRSEL	VSSQRALIS	SRAKMEEQENLLEDRL
1-1	Query_7681763	200	KELGEDAI KHAEESSE	CLEDERR		DWRSEL	VSSUQALAS	SKSKMEEQEKLLEQUL
1-1	Query_7681778	202	KRIEEEISKIGGGANK	TIEDERR	VKKLLA		VSSURALIS	SKKKMELQEN LLENKL
101	$Query_7681758$	265	KKLEEETSKYGGEASB	SVEEFRG	VKKLLA	DVRSEL	VISOBALAS	SREKMEEOENOLEECL
lcl	Ouery 7681765	294	KNLADEESEBAVDHNK	NLOEFTR	VKKLLA		VSSOESLTL	SRKKMEDOOEVLEKKL
lcl	Ouery 7681756	317	KKTAEEASKYKAEAKG	TAOELKR	VRVIIA	DVKTEL	TASOKSLAS	SGRRLDDOGVELKKOL
lcl	Ouerv 7681767	313	TELPSTVEY	.FDDLTR	VRMLLS	DVRSEL	VSSRESLIL	SRKKMEDOOAVLEHEI
lcl	Query_7681757	309	KELAQLA <mark>SK</mark> DTDDIKG	TVTDFKR	VRSLLD	AVRSEL	ISSKDNFAS	SRROIEEOTVOLOKOV
lcl	Query_7681779	303	KE LAQMA SKDKDNIKN	TISDFKR	VRSLLD	AVRSEL	IA <mark>S</mark> KEAFTF	SRKOVEDOAAOLSNKV
lcl	Query_7681783	303	KELAQMAFKDKDDIKN	TINDFKR	VRYLLD	AVRSEL	IA <mark>S</mark> KEALTF	SRKQVEDQAAQLSNQV
lcl	Query_7681772	302	KEVAQLACKDMDGIKD	TVSDFKR	VRSLLD	AVRSEL	IA <mark>S</mark> KEAFSS	SRKQIEDQAVQMQKQV
lcl	Query_7681781	304	KELAQLAFKDTDDIKG	IITDFKR	VRSLLD	AVRSEL	ISSKDAFAS	SRRQIEDQAVQLQEQV
lcl	Query_7681776	254	KELAQMA <mark>SK</mark> DKDNIKD	TIDDFKR	VRSLLD	AVRSEL	MASKEAFTF	SCRQIEDQAAQLSKQV

1c1	Query_7681755			٩
lcl	Query_7681755	364	EETEE ORKSVLSYMOSLRDAHTEVESERVKLRVVEAKNFALEREISVOKELLEDLREEL	0
lcl	Query_7681764	361	DELEEQRKSVASYMESLKNAQIEVESERVKLRVVEARNKDLERDLSVERELIKELQEEL	K
lcl	Query_7681774	361	EETEEHKGSVMTYLTTLKDAQIEVQSERAKLKVAEAQKKELERDLSMEKELMEELQELL	K
lcl	Query_7681775	357	AHLEEERKSVLSYMTSLKDAQIEVESERAKLRISEARNKELERDLSIEKELIEELHEEL	K
lcl	Query_7681759	356	VETEEQKKSLTSYMTSLKDAQVEVESERVKLRVTEARNKELERDLSMEKELVEELQNEL	N
lcl	Query_7681766	360	TEERKKNVMSYMVSLKDAQIEVESERVKLRVAQARNKELERDLSMEKELMEELQEEL	N
lcl	Query_7681770	363	VELEQRDIVMSYMSGLKDAQTEVESERAKLRIAEAQNKALERDLSMEKELIEQLQKEL	١N
lcl	Query_7681768	358	LE LE ERESVMSYMTSLKDAQMEVESEKGKLRAAEARNKELEQKLSLEKEIMEEIREEL	۱N
lcl	Query_7681762	363	TELVEQKESIAAHMASLKDVQIEVESERVKLRVAEARNKELERDLSLEKELVEELQDQL	ιT
lcl	Query_7681773	360	AE LEE QRKSVMSYLNSLKNAKIEVESERVKLRTAEARNKELERDLSMEKELVEELQKEL	ιE
lcl	Query_7681780	401	EELEEQKRSVASYMESLKNAQIEVESERVKLRVVDARNKELERDLSVERELIEELQEEL	K
1c1	Query_7681782	359	AE LEE QKTSINHYMTSLKDAQIDVESERVKLRVAESRNKELEWDLSVKKELMEELQEEL	R
lcl	Query_7681771	356	REFEE ORRSVMAYMTSLKEAQIEVESEKVKLTVAEARNRELERDLSMEKELIEELONEL	N
101	Query_/681/60	361		E
101	Query_/681/61	358	AE EE QRRSINSIIISLRDAGVEVEGERAALRVAEAQNKOLERDISLEKELVSELQKEL	
101	Query_/681///	360	EE EE ORRSVMSIMISLREAQNBVENERVUUVERERVUUVERERVUUSIERDUSIERELVEELUIEN	N
101	Query_7601703	310		
101	Query_7601778	360	LE DE VERS VINS IM I SLA BAVIN VENEAMUNET VISA ANNA ELERO LISMEA EL VESLUT ELLO ELLO EN	IN V
101	Query_7681758	325	OR TE OR D SUMSYMT SILKEV ON WETEK VKI TVA FARMKE I ED DI SIEKEI WOKI ONE'	T
101	Query 7681765	354	LE VERY LE VIENE LE VIENE SER AKLELA EARNOELKRUIS IEKDVIKELONOL	. N
101	Ouery 7681756	377	EDINE EKSULTSYMTNIKEAOLEVESERKKIRYA EARNOELEOKISKE OAM TEKLONEL	N
lcl	Ouery 7681767	365	LEFEEHBKSLSDYTRSVKDAETEVEKEBUMEBLAEGBNOEFOBDLLTEKELTDELONOL	N
lcl	Ouerv 7681757	369	OETKDORVLLMSYTODLEAAOLEIOGKTKDLNAAOSRCHELELOLLKEMEKVESLEAEL	Т
lcl	Ouerv 7681779	363	ÕETTDÕKALIISYTRNLEAAÕLEIRGKTTELNAAÕSRCSELESÕLLEETKKVESLEGML	Т
lcl	Query_7681783	363	QELTDOKALIISYTRNLEAAQLEIQGKSNELSTVQSRCSELESQLLEETEKVEFLEAML	Т
lcl	Query_7681772	362	QE SGORLLLSSFNONLEAARLEIQGKAKELNAAOSRCHELESLLQEKEKVESLEAVL	Т
lcl	Query_7681781	364	QE EED ORVLLMSYTHDLE AA OLE I OGKT QE LNYA OSRCHELES OLLKEMEKVES LEAEL	Т
lcl	Query_7681776	314	QETADQKALLI <u>SY</u> TQNLEVAQLEIQGKTNELSGVQSRCSELESQLLEKMKKVESLEAML	T
	-			

TGT Query_/081//0	514	QETADQAALLISIIQALEVAQLTIGAIAELSGVQSKCSELESQLLEKMAAVESHEAMII
1c1 Ouerv 7681755		0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
/2		
lcl Query_7681755	424	KEKPLLELAMHDISVIQDELYKKANAFQVSQNLLQEKESSLVEAKLEIQHLKSEQASLEL
1c1 Query_7681764	421	KEKYSLQLAIQDASFLRKQLGKKHTEFVEMSNVLQNKEVELVEAKLEIQHLKSERASLQL
1c1 Query_7681774	421	KERYSLHQAINGISSLQKKLDKKNADFGKMRDLLQVKESEMVEAKLEIQHLKSEQDSLKL
lc1 Query_7681775	417	KEKSSLKQAMEEMSSLREELEQKNTEFGEIHGLIQDKESELVEAKLEIQHLKSEQASLQL
1c1 Query_7681759	416	KEKYSLQQAIDEVSSLQEELGRKNTEFGETENLLRVKESDLVEAKLEIQNLKSKQASLQL
1c1 Query_7681766	420	KERTSLQQAIQDMSLLQEELDRKNTEFERMNNILQVKESESVEAKLEIQHLRSEQAALQL
1c1 Query_7681770	423	KERSSMEQAMQDISLLQEELGQKTFEFEEAQNLLQVKESTLVEARLQIQHLKSEQASIQL
1c1 Query_7681768	418	KEKSSLEQMVQEMSYLQQELAAKNTEFGEMHDILQFKESELVEAKLEIQHLKSVQCSLQL
1c1 Query_7681762	423	K EKAS LQQA I QDKS LLQTE LN RRNSEFQETQDL LRLKEGELVEAKLEI QHLHSQQTALOL
1c1 Query_7681773	420	KEKSSLQQEIEKTSFLQQELLQKNIEFGEMQHLLQAKESDLVEAKLDIQNLKSEQASLQL
1c1 Query_7681780	461	KEKSSLQQAIQDVSFLRONLEQKNAEFGEMSNVLQSKEADLVEAKLEIQHLKSERASLQL
1c1 Query_7681782	419	RERSSLOOVIQETSFLOKELDOKTTEFGELHNLLOVKESELVEARLEIOHLKSEOVSLOL
Icl Query_7681771	416	N EKSSLLAALKEKSTLQEELDHKRAEFGETQDLLQVKESELVDARLEIQHLKSECASLQL
1c1 Query_7681760	421	REKSFLQQATEEKSLLQNELEHKRIEFEKTHKLLQDKASALVEAKLEIQHLKSKQVSLQL
Ic1 Query_7681761	418	K ER LS LK EA I VEISTLRED ID RKSAAFE OS OSF LKSKESEL VEARLEI OHLKSE OAS LLL
Icl Query_7681777	420	IKKSSLHVAINEKSALQEELDCKSAEFGETQNLLQVKESELVDARLEIQHLKSQCASLQL
1c1 Query_7681763	378	KERTSLEQAVKEVAFLOEELEKKNAEFRETSAVLOVRESELVDAKLEIORLKSEKASLOG
Ici Query_/681//8	420	I KKSSLYVAINEKSALQEELD KKSAEFGEI QNLLQVIESELVDAKLEI QHLKSQCASLQL
Ici Query_/681/69	413	KEKAS LEQAVQEMALLQEELDIKSAEF KEKSALLDVKESELVDAKLQIQELKIEKASLQA
1c1 Query_/681/58	385	I RKPS LQAAMIEKSALQEELY RKS IEFGET ONLLOVKESELVDARLMIRHLKSECSELQI
Ici Query_/681/65	414	E EKKSLIQAIEELSVIRDELDRKISEYKSMRNALESKESQIVEARLEIQULKSKQAFLEV
1c1 Query_/681/56	437	KEKLSLEKIIKDVGSLIALLEQKISDLDNILKIKVKESELVSAKSEIQLLKSDHESIQL
1c1 Query_/681/6/	425	SERDSLERKANEEILEIKDELNKRNLEFSEIQNKLESKEAELVDAKIEIOSLKSERVCLEV
1c1 Query_/681/5/	429	KEREN LEOKIEOVDELOKELVOKENECONSOKLVKIKERELEAKHEVOD KSKVD SIOL
1c1 Query_/081//9	423	REFERENCE VILLOFF IV ORE ND IF NS ORL VE INFILE I FLIFARHE VEDMALK VDS I OF
1c1/Query_/081/83	423	REALING VIEWEL OVER VORENDIENS VE VETVERENT HEN VETVER VED
1al Query_7681772	422	KEKES LEEKIKEVELUQAALV QAENE ASNS LALVEIKESELLEAKNEV QDMASKVESIQI
1c1 Query_/081/81	424	K DE CLEARTER VGELQARELV QAENECIASQELVA VAEFELLEARQEVQDMKLKVESIQL
TCT Query_7681776	574	KEKES DEQKIKEVDLLQEE DAQKENEYFNSQKLVDT <u>KESELLEV</u> RHEVEDMKLKVD <u>SIQF</u>

				α6								
1c1	Query_7681755		عفقعفعف	لععفه	مععععه	ىقفق	يفعفه	لعععه	فععف	فعفعف	يعفف	عفعفعفع
lcl	Query_7681755	484	LLQEKDEEL	AEARNH	KLGEVNQ	EVTE	KALMI	SREDQ	LMEAT	EMLKEK	DVHLH	HRIEGELGS
lcl	Query_7681764	481	ILEEKDQEL	SDAKKI	NLEQLNQ	EIAEI	KMLMS	SKENQ	LIQAT	ALLKEK	DEYAJ	LKVQDELND
lcl	Query_7681774	481	ILDEKDLEL	LNARHH	KLEEVNN	EIAEI	KMLLN	SKEDQ	LIQAT	TMLKEK	DEHVN	NTMQNELND
lcl	Query_7681775	477	VLEGKDRQL	LSAKKH	KLEEVDQ	EIAE	KMLLS	SKEDQ	LIQAT	NMLKEK	EEHV(2VMQDELNE
lcl	Query_7681759	476	ILEEKDFEL	SNARQI	ЧLЕЕLNN	EVREI	KMIMS	SREEQ	LVQAM	DTLQEK	DEHVJ	LILQNELDG
lcl	Query_7681766	480	ILEEKDMEL	LNARKH	KLEGLNQ	EIGEI	KMLMN	SREDQ	LILAT	TMLKEK	DDHV	LMKDELND
lcl	Query_7681770	483	LLEEKDMDL	YDAQKI	NLEELSK	EVAEI	KMLMS	SKEEQ	LIQAT	NMLKEK	DMHV	2MMQHQLDD
lcl	Query_7681768	478	LSEEKDLQL	LDAKKI	KLEELDQ	EVAEI	KMLLS	SKEDQ	LIQAT	NMLMEK	EERV	2MMQDELND
lcl	Query_7681762	483	MLEEKDVEL	LNTKKI	KLEELNS	ETEEI	KRLMN	IGREDQ	LIQAT	NLLKDK	EEHV	QNMQSELSD
lcl	Query_7681773	480	ILEDKDLQL	FDARKI	NLDEVNQ	EVAEI	RMLMS	SKEQQ	LVQAT	TMIKEK	EEHVÓ	QVMQDELNN
lcl	Query_7681780	521	ILEEKDLEL	SNARKI	NLEQVNQ	EIVEI	KMLMS	SRENQ	LIQAA	ALLKEK	DEHVÓ	2KVQDELND
lcl	Query_7681782	479	ILKERDLEL	FNAQKI	KLEEVNQ	EVSEI	KMLMN	NREDQ	LMQAT	TLLKEK	EEHLI	LIMQHELND
lcl	Query_7681771	476	MLEEKDKEL	LDSRKN	MLDELNQ	DIAE	RVLMS	SQEVQ	LIQAT	TMLKEK	EESMÇ	2 T M Q D E L N D
lcl	Query_7681760	481	LLEEKDLEI	LDAQKI	KIQNL <mark>N</mark> Q	EIIEI	QTLMS	SKEAQ	LDQTT	AMLKEK	DERVE	ETMQNELND
lcl	Query_7681761	478	ILEEKDLEL	SNAKTN	ЧLEEVNK	DIDE	KRILF	SREEE	LTKAS	SMLKEK	DEHVÇ	2 T I E H D L S N
lcl	Query_7681777	480	MLEEKDKEL	LDSRK	ΓVDELNQ	EIAEI	RVNMN	SQEQQ	LIQAT	SMLKEK	ΕΕSΜΟ	QIMQLELND
lcl	Query_7681763	438	ILEEKDLEL	SNAKKI	MLGEINQ	DISD	KMLMN	SKETQ	FIEAT	NMLREK	DEQLH	KMIQNK
lcl	Query_7681778	480	MLEEKNKEL	LDSRK	「LDELNQ	DIAE	RVLMN	SQEQQ	LIQAT	SMLKEK	ЕЕFМÇ	QIMQLELND
lcl	Query_7681769	473	LLEEKDLEL	SSARKI	ЧLVELNQ	DISD	KMLMN	DKETQ	LIEAT	NMLREK	DEHVE	KVIQNKLNN
lcl	Query_7681758	445	MLEEKDKEL	LDSRK	ΓLΕ <mark>Ε</mark> LNQ	EIDEI	RVLVS	SQELQ	LIQAT	SVLKEK	ΕΕSΜΩ	QIIQLELND
lcl	Query_7681765	474	MLMEKESEL	SQSQEN	MLSEVNQ	DILN	KALLS	DRETE	LTETT	TMLKEK	DKLVI	EKIQHDRNE
lcl	Query_7681756	497	LLKEKDVEL	SLANKI	NLEDLDK	EVRD	KKLMF	EKEEQ	LIQAT	LGLQEK	DQLVE	EMMRLEIDD
lcl	Query_7681767	485	MLNEKESEL	SDAREI	ILDEVNR	DIMNI	KMLLG				F	RKIE
lcl	Query_7681757	489	AVQEKDSEL	SDTQSI	RLTEVSG	EVVEI	QQLLN	SKDDQ	LVQVR	TELHDK	EQYIE	ESMQSELES
lcl	Query_7681779	483	AVQEKDLEL	LETQRE	KLDEVNS	EVVE	QQLIN	ISKEDQ	LVQVR	TELQDK	ЕНСІÇ	2 L M Q D E L D K
lcl	Query_7681783	483	AVREKDLEL	LEAQRE	KLDEVNS	EVVEI	QQLIN	SKEDQ	LVQVR	TELQDK	EQCIÇ	2 L M Q D E L D K
lcl	Query_7681772	482	AVQEKDSEL	SETQRI	RLAEVNS	EVVEI	KQLLE	SKEDQ	LVQVR	TELQDK	EQHIÇ	2TLQNKLDS
1c1	Query_7681781	484	AVQEKDSEL	SDTQSI	RLTEVSS	EIVE	QQLLN	SKKDQ	LVQAR	TELHDK	EQHIE	ETLESELDS
1c1	Query_7681776	434	AVHKKDLEL	LETQRE	RLDEVNN	DVVE	QQMIS	TKEDQ	LVQVQ	PLS	PICIN	NITNWRLTS

1c1	Query_7681755		200000000000000
lcl	Query_7681755	544	SKLKVTEAEMVVERIAELTNRLLMSTTN.GQNQNAMRINNEISIDSMQQPLEK
lcl	Query_7681764	541	TKMKFSEAETVIERIAELTNRLVISVKD.ED.NNVLRPV.DDVSSELMHQLVDR
lcl	Query_7681774	541	TKLKYSEAETVVGRIVELTNKLVISVKD.DD.SNAPRMF.DDMGQDLLQQLLEN
lcl	Query_7681775	537	TKMKISEAETVVERIVELTNKLVISIKD.ED.HNAFAPS.DSTSLDLVQQPLDR
lcl	Query_7681759	536	TKLKVSEAETVVEQIVDLTHKLVISNKN.DE.SSTSMPT.DDMGLELMQQGLDK
lcl	Query_7681766	540	VKLKFSEANTAVQQIVELTNKLAMSIKD.EN.STAVSAF.DDMGHDFLQQLLEK
lcl	Query_7681770	543	TKLKYSEAATVVERIVELTNKLVISVRD.EE.DYALDTL.SKEILLPEMEDKFLQHILEK
lcl	Query_7681768	538	ARLKIS <mark>EAE</mark> TVVERIVDLTNKLVITVKD.ED.YNAVRPS.GSMDLELIRQPLDK
lcl	Query_7681762	543	TKWRCTEAESVVERIVELTNRLVIPLED.KE.YDVAKFLGVGGLGSQQLITL
lcl	Query_7681773	540	TRVKVSEAESVVERIVELTNELVISIKD.QNELRQS.NNMTLEFFQQPLDE
lcl	Query_7681780	581	TKIKFS <mark>EAE</mark> TVIERIAELTNRLVS <mark>S</mark> AKD.ED.NNVLRPV.DDVSHELMHQLVDR
lcl	Query_7681782	539	TKLKFSEAESVVERIVDLTNKLVICTKD.EE.CTATSPF.DDMGQNLLHQLFEK
lcl	Query_7681771	536	TKMKYS <mark>EAE</mark> TVVERIVDLTNKLVISVKD.DV.LSPLSHA.DETWSSQLVEK
lcl	Query_7681760	541	TKLKISEAEAAVEHIVDLTNKLVISIKDGDE.YDVLKLNENLSLNLQQQLFKK
lcl	Query_7681761	538	AKSRFTEAEMVVEKIVDLTKEAVLSFDD.EEGYHALGPLDQNNDSLTPSWLDG
lcl	Query_7681777	540	TKMKYS <mark>EAE</mark> TVVEHMVDLTNKLVISVKD.DV.LSPLSHT.DEMWSSQLVEK
lcl	Query_7681763	495	
lcl	Query_7681778	540	TKKKYLEAETVVEQMVDLTNKLVISVKD.DV.LSSLSHT.DEMWSSQLMEK
lcl	Query_7681769	533	TSLKAF <mark>EAE</mark> TVVGRVLDLTNKLVASIKN.ED.INSSRPL.NELGDQLMMPLSED
lcl	Query_7681758	505	AKMKYS <mark>EAE</mark> TVVEHIVDLTNKLVIDM.LSTLSHV.DGMWPSQLVGK
lcl	Query_7681765	534	SNLKYV <mark>EATTVV</mark> ERIFELTNKVVS <mark>S</mark> IEHEGR
lcl	Query_7681756	557	TKLKYL <mark>EAAAVV</mark> GRIMQLTNILVNTARE.ETWNLESHGSLLSGVEG
lcl	Query_7681767	520	TGIDYS <mark>EAKNVMERIFELTN</mark> KVDD <mark>S</mark> IVKPES
lcl	Query_7681757	549	IRFRCSQAESVLRRMAELTGDLASSVKA.GE.MDIYALLDDEISSTSTVL
lcl	Query_7681779	543	MRLGRSQAHSVVQKIVELTGNLIGSVES.EE.LDIYNLLDDEILSTSTAL
lcl	Query_7681783	543	MRLGRSQAESVVQKIVELTSNLIGSVKG.EE.FNIYNLLDDEILSTSTAL
lcl	Query_7681772	542	MKFSCSQAESVVQKIAELTGNLASSVEG.EE.MDIYALLDDEISSTGTAL
lcl	Query_7681781	544	IRLRCSQAESMVQRMAELTGDLASSVKA.GE.MDIYTLLDDEISSTSTAL
lcl	Query_7681776	491	TC

							0	.7
1c1	Query_7681755		فقفقه	·llll	مععععع	مقمعموهم	ممممممممم	<u>2222222222222222222222222222222222222</u>
lcl	Query_7681755	596	PHDDYGMEN	. KRLVME	LSFTREN	LRMKEMEV	LAVORALTFKDE	EINVVMGRLEAKEOELK
lcl	Query_7681764	592	PFSDFGLQK	. KQLETE	LRFTKES	LKDKEMEV	LAAQRALAIKDE	ELKMVLGRLEAREKELQ
lcl	Query_7681774	592	PADDFRLQI	. KQLETE	LELARDS	LRTKEMEV	LAFQRALTIKDE	ELKMVLGRLDAKEKEVK
lcl	Query_7681775	588	PGDYFRLQK	. EQLENE	LSLTRER	LRMKEMEV	LASQKALTIKDE	ELKAVLGKLDAREKELK
lcl	Query_7681759	587	GNDNFRLQT	. KQLEIE	LKFAREN	LRMKEMEV	LAAKRALTVKDE	ELKTVLGRLDAKEKELK
lcl	Query_7681766	591	PTDGFSLQK	. RQVETE	LKLTRES	LRMKEIEV	LAVQRALTLKDE	ELRTVIGRLDAREKELK
lcl	Query_7681770	600	PTGDTALKN	.KQLEAE	LELARES	LRIKEMEV	LAGERALIVKEE	ELKAVLDRLDAREKELK
lcl	Query_7681768	589	TSDDFSLQK	. KQLEGE	LKLTRES	LRMKEMEV	L A A Q R S L T I K D E	ELKVVLGRLDSKEKELK
lcl	Query_7681762	593	PTDDFRWKN	. KQLETE	IDLTREN	LRKKEMEL	L D A Q <mark>R A L</mark> T L <mark>K</mark> D E	ELKSVLIRLGEREKDLQ
lcl	Query_7681773	589	LSDDFRLQK	. KQYETE	LKFSRES	LRVKEMEV	L A A K <mark>R A L</mark> A I <mark>K</mark> D E	ELKTVLERLDTKEKELR
lcl	Query_7681780	632	P.NDFRLQK	. KQLETE	LKSTKES	LKVKEMEV	L A A Q <mark>R A L</mark> T I <mark>K</mark> D E	ELKMVLGRLEAREKEVQ
lcl	Query_7681782	590	PTDDFKRQE	. KRLETE	LELTRES	LRTKELEV	LAAQRALTIKDE	ELKIALERLDAREKELR
lcl	Query_7681771	584	STDTFRWQK	. NQLENE	LELTRES	LRGREMEA	LAAQRALKLKEE	ELKMVRQKLDDREEEIN
lcl	Query_7681760	593	PTDNIRLQK	. KQLETE	LELTKES	LRRKEMEI	LAAERALTVKDE	ELKTVQERLDGKEKEFE
lcl	Query_7681761	590	FGDSFKWQK	. KQLEAE	LVFTRES	LKTKEMEI	L A A Q K D L T I K D E	ELKMVIRKLEAKEKEIT
lcl	Query_7681777	588	PTDAFRWHK	. NQLENE	LELTRES	LRSREMDS	LAAQRALKLKEÇ	ELKIVRQKLNDREEEIN
lcl	Query_7681763	533	PANELRWQQ	.KRLENE	LELAKVT	LKEKEMEV	LAAQRALTIKDE	. E L K M T L A R L D S K E E E L K
lcl	Query_7681778	588	PTDTFRWHK	. NHLENE	LELTRES	LRSREMDS	LAAQRALKLKEÇ	ELKIVRQKLNDREEEIN
lcl	Query_7681769	584	PTSELSWQQ	.KQLENV	LEL		RALTIKDE	. E L KMT L A R L D A K E E E L R
lcl	Query_7681758	549	PTDPFRWQK	.NQL.NE	LELTRES	LRSRERDV	LAAQRAVKLKGQ	ELKMVHQKLIAKEEEIN
lcl	Query_7681765	565	ILSK	EKQLETE	LDLIMET	LRSREMEV	LQTRRALTVKEN	ELKMALDKLNERENEMK
TCT	Query_7681756	602	SELEQMRDR	HKETEIE	LEMTKAR	LRHTESEY.	LSAQRALSVKEE	ELQAILNEWDTREKELE
TCT	Query_7681767	551		. KRVETE	LDVIRGT	LRLREFEV	LRSRREIMIKED	KVKSVLEKLDERENEMA
TCT	Query_/681/5/	59/	ESNLHK	HNQLEAD	IEMLRES		RAAHEALDAKDQ	ELKAVVGKWDFKEKELD
TCT	Query_/681//9	591	ESNLHK	HSQLKAD	IDMLKES	LREKDMDL	SAAYKALDAKDE	ELKAVVGRLDVRDKELD
TCT	Query_/681/83	291	EISLHK	HNQLEAD	IDMLKES	LRQKDMDL	TAAYKALDAKDE	ELKAVVGRLDVRDKELD
101	Query_/081//2	590	KSNLHK	INOLEAD	TEMIDEC	L D U K D M E I	RAAHLALDAKDQ	ELNAVMKKWUVKE.EVU
101	Query_/081/81	592	ESNLHK	пиуггАD	TEMTKEC	ық пқрыд.	к <mark>ч</mark> чн гчг ракрС	ETVAALKVMDAVEKETK
TCT	Query_/081//6		• • • • • • • • •	· · [• • •] • [•	<u></u>	<u></u> <u></u> l	<u></u> <u></u> <u>.</u>	······································

																		α8			
1c1	Query_7681755		عفعفعف	مععه	يعف	لععف	ععه	l		l	ىق	عف	يوو	٥٩	مععه	l.	معف	عع	ععف	re	فعف
lcl	Query_7681755	655	KLKEETIN	DSED	LKV	YAL	AQE	RVC	EKI	Г М <mark>С</mark>	DLI	AIF	ML	QL	EAAN	L.	EVE	AAT	SAI	QK	Lak
lcl	Query_7681764	651	RLKEEMTE	DAND	MKK	YAL	AQE	RIC	EKS	SIG	DL	A I E	ΕKL	QL	EAAÇ	ь.	EVE.	AAT	SAI	QK	LAE
lcl	Query_7681774	651	KMKEEA.E	DAND	LRK	YAL	AQE	RLC	EKS	SIG	DL	A I E	ΕKL	QI	EAAÇ	ь.	EVE.	AAT	NAI	HK	LAE
lcl	Query_7681775	647	GLKDEMIE	DANC	LKK	YTL	AQE	RIC	EKS	SIG	ELI	A I E	ΕKL	QL	EAAÇ	ь.	EVE	AAI	SAI	LK	LVE
lcl	Query_7681759	646	KLEETV.E	DANC	LRK	YAL	AQE	RFC	EKS	S V <mark>G</mark>	DLI	A I E	RL	QL	EAAÇ	ь.	EVE	AAI	SAI	QK	LΤΕ
lcl	Query_7681766	650	MLKEELNE	DANC	PRK	YAL	AQE	RIC	DGS	SIG	DLI	ALE	ΕKL	QL	EAAÇ	ь.	EVE	AAI	CAI	SK	LID
lcl	Query_7681770	659	RMKEE.VE	DADG	LKK	YAL	AQE	RIC	EKS	SIG	ELI	A I E	ΕKL	QL	EATÇ	ь.	. E V E	AAI	SAI	RK	LAD
lcl	Query_7681768	648	RIKDEMLE	DAND	LNK	YSL	AQE	RIC	EKS	SIG	DWZ	A I E	ΕKL	QL	EAAÇ	ь.	. E V E	AAI	CAI	JQK:	LAG
lcl	Query_7681762	652	RLKEEMIN	DASD	LRK	HEL	AQQ	RVC	EKS	5 M <mark>G</mark>	DL	A I E	ΕKL	QL	EAAÇ	ь.	. E V E	AAI	SAI	EK	LVE
lcl	Query_7681773	648	KLKEEAVE	DAND	LRK	YSL	AQE	RIC	ESS	S V <mark>G</mark>	DLI	A I E	ΣKL	ΚL	EAAÇ	L.	. E V E	AAI	SAL	JQK:	LAE
lcl	Query_7681780	690	RLKEEMVE	DAND	LKK <mark>I</mark>	YAL	AQE	RIC	EIS	SIG	DLI	A I E	ΣKL	QL	EAAÇ	L.	.EIE	AAI	SAL	JQK:	LAE
lcl	Query_7681782	649	RMKEETME	DANH	ILKN	YAL	AQE	RIC	EKS	S V G	DLI	A I E	ΕKL	QL	EAAÇ	ь.	. E V E	AAT	SAI	HK	LAE
lcl	Query_7681771	643	KMKEMT.R	DADD	LMQ <mark>I</mark>	YAL	AQE	RIC	EKS	STG	DLI	A I E	ΣKL	QL	ЕТАС	L.	. E V E	AAI	SAL	JQK:	LAE
lcl	Query_7681760	652	KMKEEMDE	EGKH	ίLRΕÇ	YTL	AQD	NVC	• • •	G	DLI	AIE	RL	QFI	EAAÇ	L.	. E V E	AAI	SAL	JQK:	LTD
lcl	Query_7681761	649	EMKG	DKDG	; I K Q <mark>I</mark>	YAL	AQE	RIC	DKS	S V <mark>G</mark>	DLI	AIE	ΣKL	QFI	EVAÇ	L.	. E V E	AAI	SAL	JQK	ΙAΕ
lcl	Query_7681777	647	KMKNMT.R	DADG	SPRQ S	YVL	AQE	RTC	EKS	STG	DLI	AVE	ΕKL	QFI	EGAÇ	ь.	- E V E	AAI	TAI	JQK:	LAE
lcl	Query_7681763	592	KVREEVTE	DSND	LKR	YAL	AQE	RIC	EKS	δL <mark>G</mark>	DLI	A I E	ΣKL	QL	EAAÇ	L.	. E V E	AAI	'NAI	JQK:	LAE
lcl	Query_7681778	647	KMKEMT.Q	DADG	;VRQ <mark>I</mark>	YAL	AQE	RTC	EKS	S T <mark>G</mark>	YLI	A V E	ΕKL	QFI	ERAÇ	ь.	. E V E	AAT	SAI	RK	LAE
lcl	Query_7681769	627	KAKDMATE	DAND	HKM	YAM	ΤQΕ	RIF	EKI	ΓMD	DL	A I E	ΕKL	QL	EAAÇ	LEI	DEVE	AAI	STL	JQK:	LAE
lcl	Query_7681758	607	KMKEKT.R	DRDD	LKQ <mark>I</mark>	NAL	AKE	RTC	EKS	SIG	DPI	A I E	ΣKL	QL	ККТÇ	L.	. EVN	AAI	SAL	JQK:	LVE
lcl	Query_7681765	620	EMKQELTK	DADD	LRK	YEM	VQE	RIC	DKS	SIG	ELI	AIE	ΣKL	ΕL	EAAÇ	L.	. E V E	AAI	SAL	EK	LTE
lcl	Query_7681756	662	KLQENVQE	EVNG	JLQKE	QAI	ADD	GI	NRS	SIE	E L 1	ΓMΕ	ΕKL	QL	EAAÇ	ь.	. E V E	VAI	SAI	RN	LAN
lcl	Query_7681767	605	EMKWELSQ	DVDE	LRR	YAM	AQE	RIC	ERI	ГМ <mark>С</mark>	ELI	A I E	ΣKΜ	ΕL	EAAE	I.	.EIE	AAV	′SAI	EK	ΙMΕ
lcl	Query_7681757	654	EV.EELQK	DPIC	MKE <mark>I</mark>	PVL	SNE	TT	GSI	ΓTG	EME	ΞLF	ΚKL	QII	EAAE	v.	. EAL	AAI	TAI	KK	LAD
lcl	Query_7681779	648	KL.QELSI	DPYD	IRR	SSA	A D E	ΑTΡ	DNI	I V E	ΕLΕ	ΞLζ	2 <mark>K</mark> H	ΕII	ESVE	v.	. EAI	AAS	ТΜΙ	KK	LAN
lcl	Query_7681783	648	KL.EELSI	DPYG	TRK	SRV	ADE	ATE	DNI	[A <mark>G</mark>	EAE	ΞLζ	2 <mark>К</mark> Н	ΕM	ESVE	м.	. EAL	AAS	ΤMΙ	KK	LAD
lcl	Query_7681772	646	KL.EGFLK	PSD	IKRF	SDF	SVH				.MC	GΓζ	QΝL	Q T I	EAAE	v.	. EAL	AAI	TTI	KK [K]	LAD
lcl	Query_7681781	649	EL.EELP.	PSA	TNE	AGF	SSE	TTE	GGI	C V <mark>G</mark>	ΕMΕ	ΞLΕ	? E L	QI	D <mark>A A</mark> E	v.	. EAI	AAT	TAI	RK	LAD
lcl	Query_7681776							<u> </u>											J. <u></u>	<u>_</u> .	

				α9
1c1	Query_7681755		22222222222222	
lcl	Query_7681755	713	MSTELLTQADMSIEAD	TTHTVMPERGYSEGSNECLGEVKTE
lcl	Query_7681764	709	MSHELLIKASTSIEAD	
lcl	Query_7681774	708	MSGEFLHKASLSIEAD	AYTTILLPNGSDPSRSAAENDECLTEVTTE
lcl	Query_7681775	705	MSRELLNKANLSIMAD	ADAETDISMFLQNYSDPGISMFGNNECLKEVKTG
lcl	Query_7681759	703	MSGELLNKASLSIETD	TDNTIFPESRFDPRISVIENNECLTEVGSE
lcl	Query_7681766	708	MGRQLLNKASVSLDAD	NDASVLPQDDSDTGIDVVGHSECLTEVKYE
lcl	Query_7681770	716	MSRELLKKASFCAEVD	IDTLVFPEPGNDPRTSVIENNECLIEAQKE
lcl	Query_7681768	706	MSRELLNKANLSVEAD	DNVDMFTQSSSDSKISIFENNECLK <mark>E</mark> VKTG
lcl	Query_7681762	710	MTHEFLGTASLSIEVD	TDASGLHEELSDMTSSDENSECFTEVKMQ
lcl	Query_7681773	706	MSRELLNKASLSIEAD	AD IFMPNGSGPGLVLLENN ECFKEVKTE
lcl	Query_7681780	748	MSRELLNKASMSVEAD	
lcl	Query_7681782	707	MSCELLHNVSLSVDSE	TD TAIFLPNGFDPWLSMHENN EHFTKVKTE
lcl	Query_7681771	700	LSRELLNKASLSIEAD	YN NSLLLGDSPGTAAN VASSG ECLAEVYTE
lcl	Query_7681760	706	MSRDLLNKAGRSLEAD	IGS RSIRIQQHDDDNNG VNGIDNNNSRFNEVKVE
lcl	Query_7681761	703	MSRELLNKTGLCVELEASD	YDMSLYKKDNTEARINTINANKCSVEVQSE
lcl	Query_7681777	704	LSRDLLNKASLTIEAD	YDSSLLLVDIPETAANVSSSFECLAEVYSE
TCT	Query_7681763	650	MSRQLLNKAILSVEAD	
lcl	Query_7681778	704	FSRGLLNRASLTIEAD	YDSSLWLVDIPETAANVSSSFECLAEVYTE
lcl	Query_7681769	687	MSQQLLNKAMPSVEAD	
TCT	Query_7681758	664	LSRDLLNKGILTIESD	YDNSLLLVAHSEAAVNVTSSVQYLAKVHTE
TCT	Query_7681765	678	MSRELVRITSLLIDAD	YDLDVSTKNIPEREVPVVENGCFAELKTE
TCT	Query_7681756	720	MSQTLLKGTETHLKFDSPD	YITMAQSKMEILDEEDPETSYEPSLNESLDATESFLKTKEA
TCT	Query_/681/6/	663	MSRELLRAISVIVDAD	SDVDVSMGIERCGFEDPGFEELRME
TCT	Query_/681/5/	/11	MSKKYLRCRKAD	
TCT	Query_7681779	705	VTKEFLRNGRTD	
TCT	Query_/681/83	705	VIKKFLKSGRTD	
TCT	Query_/6817/2	694	MAKGFLKSGKTD	
TCT	Query_/681/81	105	MIKDFFKHVKAD	
TCT	Query_/681//6		<u>••</u> ••••••••••••••••••••••••••••••••••	•••••••••••••••••••••••••••••••••••••••

1c1	Query_7681755		
1c1	Ouerv 7681755	754	VVRLWSLTEKLLENAGIVAGTSTCMEGVIL
1c1	Query_7681764	755	VARLSALSEOLVKDAGIVGAHLOS
1c1	Query_7681774	754	VSRISALTDOLVKEAGIVIRAGPAGOOSCSPARRS
lcl	Query_7681775	755	VVRLSAMTEQLVKEAGVAAGVN
lcl	Query_7681759	749	VARLSVLTEQLVKEAGIVDGQIS
lcl	Query_7681766	754	VERLSALTEQLVQEAGIA
lcl	Query_7681770	762	VARISALTEQLVKEAGIV.GVADQ
lcl	Query_7681768	752	VARLSALTEQLVKEAGVVVGEA
lcl	Query_7681762	755	VTKLAALSEQLVREAGITIAAE
lcl	Query_7681773	750	VARLSSLTEQLLQDAGITVGAD
lcl	Query_7681780	794	LARLSALTEQLVKDAGIVGAQLQS
lcl	Query_7681782	753	VARLSAITDQLVQEAGVVGAAN
lcl	Query_7681771	746	MARLSALTEQLVKEAGILCPK
lcl	Query_7681760	756	VSRLSSLTEQLLKEAGIFLDAD
lcl	Query_7681761	752	VSRLLTLTQQLVVEANVTGYMSQ
lcl	Query_7681777	750	MAQLSALSEKLVKEAGILCPQ
lcl	Query_7681763	698	VARLSALSEQLVMEAGIVPAN
lcl	Query_7681778	750	MTQLSALSEKLVKEAGILCPQ
lcl	Query_7681769	735	VARLSALTEQLVMDAGLAAAS
lcl	Query_7681758	710	MARLSAMTEKLVKEAGSLCPQ
lcl	Query_7681765	723	VARLSDFTQKLIQDAGIAGNSGDVCG
lcl	Query_7681756	780	IAHLSALTDKLLEDAGIAVKD
lcl	Query_7681767	704	AARLSDFTEKLVREAGIGKDLAVVDK
lcl	Query_7681757	758	IVRLFSLTKELITDDAINDA.EER
lcl	Query_7681779	752	IVGLFSLTEELVAGAQMKDA.EEP
TCT	Query_7681783	748	IVGLESLTEELVTGAQTKDDDEEP
Tcl	Query_7681772	741	IAGLESLTEQLITEAGIDVA.HQA
Tcl	Query_7681781	752	INKTERTKÖTALDDIINDA EE
lcl	Query_7681776		· · · · · · · · · · · · · · · · · · ·